Aside from a relatively small amount of definite experimental data one's judgment is finally formed on the basis of his interpretation of the vast accumulation of material comprised in the recorded and unrecorded experience of the breeders of registered (pedigreed) livestock.

This material recorded in the books of registration far exceeds in amount and in diversity any which could possibly be obtained experimentally on the same forms of life. It must be said, however, that the discussion of it with a view to an analysis of the effects of inbreeding, though undertaken at greater or less length by a number of men including Lehndorff, von Oettingen, Bruce Low, Hoesch, Chapeaurouge, Bunsow, Strang, and others, has not led to results characterized by the precision, the definiteness, or the quality of getting at fundamentals demanded in the present state of the science of genetics.

The lack of precision and fundamental character in the studies alluded to is not primarily to be attributed to any inherent defect in the material. In the breeding of all of the domestic animals inbreeding has been practiced; in many instances to a very marked degree. Further, the manner in which the inbreeding has been done (the types of relationship-matings) exhibits a most intricate diversity, from which different types may be picked out for analysis in any reasonable quantity. The records are accurate, within their limitations,

to a high degree. Probably no experimentalist's records of descent are more accurate, considering the relative numbers involved in the two cases.

The real need, I venture to think, has been for an appropriate and valid method of pedigree analysis, which possessed generality, and could on that account be depended on to give comparable results when applied to two (or more) different pedigrees. Specifically, there seems not to have been worked out any *adequate general method of measuring quantitatively the degree of inbreeding which is exhibited in a particular pedigree.* Without such a measure it is clearly impossible to proceed far in the analysis of the kinship aspect of inbreeding.

It is the purpose of this paper to present a method for measuring and expressing numerically, in the form of coefficients, the degree of inbreeding which exists in any particular case. I shall endeavor to show that the method is (*a*) *unique*, in the sense that the values obtained in any particular instance can only be affected by the degree or amount of inbreeding which has been practiced in the line of descent under consideration, and (*b*) *general*, in the sense that it is equally applicable to all pedigrees and to all degrees and types of inbreeding.

## Preliminary Definitions

In attempting any general analysis of the problem of inbreeding from the theoretical stand-

MBL/WHOI
0 0301 0011660 4

THE MACMILLAN COMPANY
NEW YORK · BOSTON · CHICAGO · DALLAS
ATLANTA · SAN FRANCISCO

MACMILLAN & CO., LIMITED
LONDON · BOMBAY · CALCUTTA
MELBOURNE

THE MACMILLAN CO. OF CANADA, LTD.
TORONTO

575.2
P31

# MODES OF RESEARCH IN GENETICS

BY

RAYMOND PEARL

BIOLOGIST OF THE MAINE AGRICULTURAL EXPERIMENT STATION

New York
THE MACMILLAN COMPANY
1915

*All rights reserved*

Copyright, 1915,
By THE MACMILLAN COMPANY.

Set up and electrotyped. Published December, 1915.

10391

Norwood Press
J. S. Cushing Co. — Berwick & Smith Co.
Norwood, Mass., U.S.A.

# PREFACE

This book has been prepared in the hope that it may meet, in some small degree, what seems to be a rather definite need in biological literature. During the past fifteen years there has been a great movement of biological research towards the problems of organic evolution, and particularly towards that phase of those problems which constitutes the subject currently known as genetics. Experimental breeding has become the order of the day, not alone in the laboratories of zoölogy and botany, but in the agricultural experiment stations and colleges everywhere.

Research in a new field means the application of new methods. Nowhere has this been more true than in genetics. For a critical handling of many genetic problems, a thorough grounding in chemistry, physics, and mathematics as well as biology is really a necessity. In particular the widespread use of the biometric technique in biology and agriculture demands a clear understanding of certain fundamental mathematical principles. Unfortunately such a clear grasp of underlying principles is too often plainly lacking in the bio-

metric work which is being turned out in ever-increasing volume.

Demands which have come to the writer from teachers of genetics and biometry in various agricultural colleges and universities for separate copies of one of the papers reprinted in this volume, in order that they might put it in the hands of their students, have suggested that there is a rather widespread and genuine desire to examine carefully the underlying methodological bases of modern genetic science.

It is hoped that the present volume may serve as an introduction to this subject. Even though the reader may not agree with either the reasoning or the conclusions, if he is stirred up to further thought and reading about the matter, the purpose of the writer will have been served. In these days of extreme specialization and great activity in biological research far too little attention is paid to the fundamental philosophical background of science.

The writer wishes to acknowledge his indebtedness to Professor H. S. Jennings for very helpful suggestions and criticisms in regard to the problems discussed in Chapter III, and to Professor J. McKeen Cattell, for permission to reprint from *Science* and the *American Naturalist* such portions of the book as have already appeared in those journals.

# TABLE OF CONTENTS


| CHAPTER | | PAGE |
|---|---|---|
| I. | A Critical Examination of Current Modes of Research in Genetics . . . . | 1 |
| II. | Biometric Ideas and Methods in Biology. Their Significance and Limitations . | 42 |
| III. | On the Nature of Statistical Knowledge . | 73 |
| IV. | Certain Logical and Mathematical Aspects of the Problem of Inbreeding . . . | 101 |
| V. | Genetics and Breeding . . . . . | 159 |
| | Index . . . . . | 179 |

# MODES OF RESEARCH IN GENETICS

## CHAPTER I

### A CRITICAL EXAMINATION OF CURRENT MODES OF RESEARCH IN GENETICS[1]

To any one interested in the philosophical bases and methods of science there is nothing which is regarded as sacrosanct merely because traditional. In theory, of course, such a condition of affairs obtains in all scientific work, and the implied state of mind is the ordinary one of all scientific workers. But while theoretically science is free of all trammels of tradition, practically the great bulk of scientific work is, in respect of its methods, most closely bound by tradition. Orthodoxy does not more firmly rule in the vestry than in the laboratory, in actual fact. Yet every one will agree that a real and permanent step in advance in science is only taken when some one looks at an old problem and at common facts from a naïve point of view, and applies to them unorthodox

[1] The substance of this paper was read before a club in Boston on May 17, 1914. It has since been modified in certain minor particulars.

methods of research. I shall try to point out, with reference to an old problem, where it would seem that current methods have encountered definite limitations and wherein lies the special value of each. If I can succeed in doing even so much, something will have been gained, even though it is not now possible to show any results achieved by new points of view and new methods in the field to be discussed.

The problem to which attention is invited is one of the most fundamental of biology, the problem of heredity. In what is to follow I shall try to do three things.

*First:* to define the problem in strictly objective terms, free if possible from any implications which have grown out of theories of the hereditary process.

*Second:* to examine critically the four chief methods of investigation by which attempts have been made to solve the problem, endeavoring to show in how far each has failed to take us beyond a certain point in our understanding of the matter.

*Third:* to draw such conclusions from this methodological critique in regard to the most promising lines for future attack as seem warranted by the results.

## I. The Problem of Heredity

The problem of heredity can be easily defined in a *general* way in terms which are perfectly

objective. It is a matter of common observation that there is a greater or less degree of resemblance in respect of all sorts of traits or characteristics between relatives, and in particular between parent and offspring. By heredity is meant the complex of causes, not now further specified or defined, which, taken together, determines this likeness or resemblance between individuals genetically related to each other. From a purely formal logical standpoint the problem of heredity is the problem of the analysis of this complex.

The difficulties of the problem, both methodological and technical, arise from certain reasonably obvious relationships between genetically connected individuals. In order to see what these are let us attempt to list in strictly objective terms, and into broad categories, what is actually known about the relationship of two individuals standing in the genetic series as parent and offspring. For the purpose of the present analysis it is desirable that the categories in such a list shall be broad ones.

The phenomena of heredity may be divided in three essential categories. These may be listed as follows:

*A. Resemblance between adult individuals.*

This resemblance is the central observed fact of heredity. Every individual organism is different in some degree from every other one, but any particular individual is more like the individuals genetically closely related to it, than it is like

other individuals. This resemblance marks the end stage of heredity regarded as a process. It is unnecessary to cite detailed instances to prove how extraordinarily minute and specific this resemblance ordinarily is, in respect of some one or more characteristics of the individual. Such facts are matters of common knowledge.

*B. Gametogenesis. The intervention of a relatively undifferentiated stage (the germ cell) in the cycle of reproduction of the individual.*

In higher organisms ordinarily each time an individual reproduces itself it does so by means of a single cell, which separates completely from the other cells which together make the individual. After separation this cell may for a time derive its nourishment from the individual which produced it, but morphologically and physiologically the completely formed gamete is essentially a separate and independent entity, with certain limitations as to the possibility of its continued independent existence.

Gametogenesis is clearly one of the most fundamental parts of the hereditary process for two reasons: first, because the gametes are the essential links in the preservation of physical continuity in the genetic series; and second, because, since all gametes formed by the same individual are demonstrably not absolutely alike in respect of their hereditary qualities, the distribution of the different hereditary qualities amongst

the several gametes must be determined, proximately at least, even if not initially, during the process of gamete formation.

*C. Somatogenesis.* (*Development and differentiation.*)

The adult offspring, in whose adult characters we perceive a resemblance to the parents, is the result of a long and complicated process of development and growth from a single cell. This cell is itself a composite structure produced by the fusion of two cells, one derived from each of the parents. From the standpoint of heredity the most striking thing about the fertilized germ cell is that it carries the potentiality of producing a higher degree of differentiation in the individual which develops from it, than it exhibits in its own structure. The existence of such potentiality is demonstrated by the *specificity* of the developmental behavior of the fertilized egg. Under no circumstances does a hen's egg ever develop into a turkey.

While the germ cell is distinguished from other cells by its potentiality, when separated from other cells of the body and appropriately stimulated, to develop and differentiate, yet it must be recognized that this is by no means a unique property of germ cells. Early studies on regeneration and the development of isolated blastomeres, and the recent experiments on the culture *in vitro* of more highly differentiated somatic cells, show clearly enough that this property is common in

greater or less degree to nearly if not quite all living cells. The distinction between germ cells and somatic cells in this regard is clearly one of degree, not kind.

Of much greater significance from the standpoint of heredity than the potentiality for development, though this of course in itself constitutes one of the fundamental problems of biology, is the specificity of the process, at once unique and manifold. Not only does any particular hen's egg produce always a hen, but it is also a particular kind of a hen which is produced, the particularity extending to the most minute details.

So much then is, in general, to be objectively observed about heredity; namely, first the resemblance between genetically related adult individuals. Further, this resemblance is dependent upon, because inseparably connected with, the two processes of gametogenesis and somatogenesis.[1]

[1] The discussion here and throughout has as primary material objects of the reasoning sexually reproducing multicellular organisms. Fundamentally it would appear, however, that there is no essential difference, so far as the elements of the hereditary process are concerned, between such higher forms, and the protozoa or other forms reproducing asexually by fission or otherwise. Such a conclusion seems certainly justified from Jennings' studies on inheritance in Paramecium reproducing by fission. Cf. in particular, in this connection, Jennings, "Heredity, Variation, and Evolution in Protozoa. I. The Fate of New Structural Characters in Paramecium, in Connection with the Problem of the Inheritance of Acquired Characters in Unicellular Organisms." *Jour. Exper. Zool.*, Vol. V, pp. 577–632, 1908.

*D.* From these observed facts two definite inferences are not only plain, but indeed inescapable. They are:

1. That germinal substance is innately possessed of a definite and particular *specificity*, which reaches, in degree, to the order of the individual, and which finds its most obvious objective expression in the specificity of somatogenesis; and

2. That the processes of reproduction are of such a sort as to tend to maintain this specificity from generation to generation.

In the light of this summary analysis of elementary facts it seems clear that *the critical problem of inheritance is the problem of the cause; the material basis; and the maintenance of the somatogenic specificity of germinal substance.* This result is, of course, not novel. Though somewhat differently worded here, it is essentially the same conclusion as was expressed by Conklin[1] some years ago, by Jennings,[2] and still earlier by Driesch.[3]

Before proceeding to the next division of our program I should like to digress for a moment to point out that there are clearly to be dis-

[1] Conklin, E. G. "The Mechanism of Heredity." *Science*, N.S., Vol. 27, pp. 89–99, 1908.

[2] *Loc. cit.*

[3] Driesch, H. "Zwei Beweise für die Autonomie von Lebensvorgängen." *Verh. V. Internat. Zool. Cong.*, Berlin, 1902.

tinguished different orders or degrees of the somatogenic specificity of the germinal substance. This fact has led to a good deal of confusion in the use of the term "heredity." For example, it is no doubt entirely correct to say that birds have feathers because this sort of dermal covering is hereditary in the class of animals specified. But heredity in this sense means, both theoretically and practically, a very different thing than when the equally true statement is made that a Barred Plymouth Rock female inherits the barred color pattern of her feathers from her sire only, and not at all from her dam. In the first case one is dealing with a phyletic matter, in the other case with a sub-varietal. It is obvious that the degrees of germinal specificity which determine the two sorts of hereditary phenomena indicated in the example must be of widely different orders. One represents the substratal or general element of heredity, the other the more superficial or individualistic element. It is obvious that the number of distinguishably different orders of germinal specificity is as great as the number of distinguishably different orders of variability, which in turn is more or less closely reflected in the various subdivisions or stages in taxonomic classification. For practical purposes of thinking and experimentation, however, it is sufficient to distinguish two orders of germinal specificity; namely, a general and a special. It would be im-

possible practically to delimit either category precisely, and, for present purposes certainly, this is not necessary.

We may now turn to the second division of our program.

## II. Methods of Research in Genetics

### 1. *The Biometric.*

The biometric mode of attacking the problem of heredity owes its inauguration to the late Sir Francis Galton, and its great development to Professor Karl Pearson. Galton was the first to perceive the importance of dealing with heredity on a statistical basis. As Pearson has well said, the recognition of this fact was one of the "greatest services of Francis Galton to biometry." Merz[1] has pointed out that Galton in his own mind sharply separated the problem of heredity into two parts, the one having to do with the material basis of hereditary phenomena, the other with the phenomena themselves. His early experiments dealt with the first of these problems. By means of blood transfusion he hoped to test and further develop Darwin's theory of pangenesis. Certain now obvious technical difficulties of dealing experimentally with the problem in this way undoubtedly had much to do with turning Galton to the study of the other phase, towards

[1] Merz, J. T. "A History of European Thought in the Nineteenth Century," Vol. II, p. 613, 1903.

which his own inherited instinct for the statistical method in general must, in all his thinking about the matter, have strongly urged him. By attacking the problem on the statistical side Galton felt that it would no longer be necessary "to embarrass ourselves with any details of theories of heredity beyond the fact that inheritance either was particulate or acted as if it were so." [1]

What Galton specifically attempted to do in attacking the problem from the statistical side was to measure precisely the degree of resemblance which obtains, in respect of different somatic characters, between individuals related in different ways in the genetic series. His primary object was to determine the proportionate contribution of each ancestor to the individual's inherited make-up. The only practical way of doing this appeared to be to measure the degree of resemblance between individuals. This remains to-day the point of view of the biometric school of genetic study. While Pearson has introduced many refinements of method, and has taken account of various modifying factors, such as assortative mating, selection and the like, it still remains the fact that in his studies on heredity what he has primarily been concerned in doing is to measure, by the method of correlation, the degree of resemblance which exists between relatives of different sorts.

[1] Galton, F. "Natural Inheritance," p. 193, 1889.

Galton himself was not able to solve completely the purely mathematical problems involved in his study of heredity. This phase of the work was undertaken by Pearson with brilliantly successful results. He developed independently a very complete and adequate system of dealing mathematically with biological data. His results in this direction form a permanent and epoch-making contribution to the working technical equipment of science. The great value of these results for the future development of biology is not yet generally realized. Merz compares their significance for biology to that of the mathematical inventions of Fourier for physics. I should be inclined personally to rate them even higher.

Pearson has applied this technique particularly to the study of inheritance. The chief result in his hands has been the elaboration of Galton's "Law of Ancestral Inheritance." This law as originally stated by Galton was as follows:[1] "There was found an average (contribution) of one fourth from each parent and one sixteenth from each grandparent. According to this geometrical scale if continued indefinitely backward the total heritage of the child would be accounted for." Pearson[2] has put this in a more general form which does not restrict the values of the fractional

[1] "Natural Inheritance," p. 195.

[2] Pearson, K. "The Law of Ancestral Heredity." *Biometrika*, Vol. II, pp. 210–228, 1903.

regression coefficients to the values given by Galton, which were admittedly determined from meager material.

Looking critically at the biometric method of studying heredity and the results which it has achieved, there are two points which seem particularly to require discussion. First, it appears clear that this method attacks the problem at what might be called its lowest level. In other words, the biometric method deals only with the element of the problem which was designated as *A* in our preceding analysis (cf. p. 3). As it has been developed in the hands of its chief devotees the biometric method of attack fails entirely to recognize any real or essential significance to the elements *B*, *C*, and *D* of our analysis; that is to gametogenesis, somatogenesis, or the specificity of the germinal substance. Herein lies, in the opinion of biologists generally, I think, the chief weakness of the results. These results constitute essentially no more than a rather precise description of the most superficial external features of the phenomena of heredity. Except only in the simplest of events (and then not directly) a description, however minute, of those events cannot give the slightest real evidence as to their cause. A scientific visitor from another planet might describe minutely the curious polymorphism to be observed amongst human beings in respect of the amount and distribution of the

hirsute ornamentation of their countenances, and tabulate vast statistics on the point. Yet at the end he would be no nearer knowing the cause of a beard on Mr. A's face, and the absence of one on Mr. B's, than when he began. The law of ancestral inheritance seems to be philosophically in much the same case.

It should be noted that in what has just been said I am referring to "description" as a working *method* of research for the acquisition of new knowledge, not to "description" as a general philosophical category constituting the means of expression of the results of experience. There is obviously a real distinction here. As a method of science, description has always held an honorable place in biology. It is indispensable, but not complete or final. The function of the descriptive method, *qua* method, in biology would appear to be essentially only to establish the basis or foundation for the application of the experimental method, which is the only strictly objective *analytic* tool that science has.[1]

Description as a general mode of expression of experience takes a unique place in Pearson's philosophy of science. His position has always been that all science is nothing but description and never can be anything else. This is a

[1] Cf. the recent valuable paper by Jennings: "Causes and Determiners in Radical Experimental Analysis." Amer. Nat., Vol. XLVII, pp. 349–360, 1913.

defensible position and has been so universally adopted as to have become almost a commonplace of scientific thought.[1] It, however, has nothing to do with the value or meaning of description as a method of research.

To return to the discussion of the biometric method in genetics, the second point to be made concerns the reasoning involved in the use of the correlation method in the study of hereditary resemblances. It has repeatedly been the boast of the biometric writers on this subject that their results were absolutely free from any biological theories. To this some of the more wicked critics have retorted that their results were also quite free from any biological significance. Such a criticism is not true, but it has in it an element of verity. The reason why it has I pointed out some years ago in another place,[2] but for the sake of the argument it may be sketched here. The essential point is the difficulty of interpreting any correlation coefficient in the terms of the causes which led to its existence. Baldly stated the argument of the current biometric method of studying heredity is this: Since by the method of correlation the degree of resemblance between

[1] One should note that the completely restrictive element of this view is by no means universally accepted. Cf. Royce's Introduction to Enriques' "Problems of Science," Open Court Pub. Co., 1914.

[2] Cf. *infra*, chap. II. "Biometric Ideas and Methods in Biology," pp. 42–72.

parent and offspring can be measured, and since heredity is a cause of the resemblance between parent and offspring, therefore the correlation method measures the degree or intensity of inheritance. This argument seems sound but it is really specious. For heredity is *not the sole cause* which can lead *statistically* to a significant correlation between parent and offspring. Anything whatsoever which tends to bring about local or group differentiation within the sample included in the table will tend to produce the same result, altogether independently of any genetic relationship or the absence of it. Such possible differentiating factors are so numerous and so difficult of detection in many cases, as to make exceedingly dangerous any assumption that, in a particular case, the coefficient of correlation is a measure of heredity.

The essential difficulty here is just as real whether one goes to the length of calculating correlation coefficients or not. The logical fallacy involved is particularly insidious in the case of attempts to study by statistical methods alone the problem of the effects of selecting fluctuating variations. Starting with heterogeneous material, as all such studies except those on self-fertilized plants or protozoa have started, if one does not keep an exact pedigree record of every single individual[1]

[1] Furthermore, it is not enough merely to keep the pedigrees. The individual pedigree-line-of-descent must be the *unit of analysis* rather

born in the experiment it becomes an absolute logical impossibility ever to determine critically whether any apparent result of the selection is due, on the one hand, to a sorting out of preëxistent hereditary differences, or, on the other hand, to a cumulative change in the hereditary determiners themselves. This proposition is not only absolutely unassailable from the logical standpoint, but that it is also practically demonstrable, has, I think, been proved by Surface's [1] analysis of the Illinois corn work, East's and Hayes' [2] selection experiments with tobacco, the

than the correlation table, which is a device admirably calculated under certain conditions — and those just the conditions which obtain in the study of selection problems — to obscure to the point of complete concealment facts which are perfectly clear and evident so long as the individual is made the unit of study. To lump material into a correlation table, with complete loss of any opportunity thereafter to get any useful knowledge about the individual, is only justified when from the nature of the material little or nothing is or can be known about the individual case beyond the fact of its occurrence. Then may we properly turn to the statistical method as a last resort in the search for knowledge. But surely in fully pedigreed material we are a very long way from knowing nothing about the individual. On the contrary, we know a most important thing about *each* individual; namely, its own particular ancestry.

The logical points regarding the statistical method here touched upon are more fully discussed farther on. (See p. 69.)

[1] Surface, F. M. "The Result of Selecting Fluctuating Variations." Data from the Illinois Corn Breeding Experiments. IV[e] Cong. Int. de Genetique, Paris (1911), pp. 222–237 (1913).

[2] East, E. M., and Hayes, H. K. "A Genetic Analysis of the Changes Produced by Selection in Experiments with Tobacco." *Amer. Nat.*, Vol. XLVIII, pp. 5–48, 1914.

Hagedoorns'[1] discussion of the results of Castle and Phillips'[2] on piebald rats, Shull's[3] analysis of the inheritance of capsule in *Bursa bursa-pastoris*, and finally my own[4] studies on fecundity in fowls.

It should be clearly understood that the above criticisms are in no wise directed against biometric methods *per se*, as weapons of research in attacking the problems of genetics. On the contrary, as Galton perceived, statistical methods are indispensable in this field. These methods are indispensable, in this particular case, because a multitude of separate and distinct causal factors discretely distributed in respect of their action, are concerned in the determination of the make-up of the adult organism. Since the locus of action of all of these factors is in each case the *individual*, it is impossible, generally speaking, to study the action of any one factor free of the influence of

[1] Hagedoorn, A. L. and A. C. "Studies on Variation and Selection." *Zeitschr. f. ind. Abst.- und Vererbungslehre*, Bd. XI, pp. 145–183, 1914.

[2] Castle, W. E., and Phillips, J. C. "Piebald Rats and Selection. An Experimental Test of the Effectiveness of Selection and of the Theory of Gametic Purity in Mendelian Crosses." Carnegie Institution Publication No. 195, pp. 1–56, 3 plates, 1914.

[3] Shull, G. H. "Duplicate Genes for Capsule-Form in *Bursa bursa-pastoris*." *Zeitschr. f. ind. Abst.- und Vererbungslehre*, Bd. XII, pp. 97–149, 1914.

[4] Pearl, R. "The Mode of Inheritance of Fecundity in the Domestic Fowl." *Jour. Exper. Zool.*, Vol. 13, pp. 153–268, 1912. Cf. also earlier papers listed in the bibliography of the paper here cited.

all others. This directly implies a necessity for the application of statistical methods. It means that it will always be necessary in studying heredity to use as large a number of individuals as possible and so determine average effects of the different causal factors. We have, of course, in genetics a special case of the general proposition, more fully developed in a later chapter (cf. *infra*, pp. 54–61), that in practice the statistical method is a logically necessary adjunct to the experimental method generally.

What we have to distinguish clearly between is, on the one hand, the biometric school of genetic study, which stands off by itself in sharp distinction to other modes of attacking the problems of heredity, and which is criticized in what has preceded; and, on the other hand, the statistical method as a general method of science, which *as such* is indispensable, as well in the study of genetics as elsewhere.

To summarize: it is believed that our analysis has shown that the purely statistical mode of attacking the problem of heredity, as it has actually been developed, finds its chief limitations in that, first, it deals only with the most superficial aspects of the problem, and second, that the particular method pursued is based upon a logically and biologically unsound assumption. At the same time statistical methods in general are found to be essential in dealing with the problems.

2. *The Mendelian.*

From a methodological standpoint the Mendelian mode of studying inheritance is a statistical method. The fire-eating Mendelist, if any more such exist, to whom biometry is anathema, may object to this statement, but a little consideration will show it to be true. In actual fact, the most essential methodological difference between the biometric and the Mendelian methods is one which has hitherto been quite generally overlooked, so far as I am aware. It is found merely in the fact that *the biometric method studies the ancestry of the individual, while the Mendelian method studies the individual's progeny.* One goes backward on the pedigree; the other goes forward. The network of descent may be likened to two pencils of light rays both of which focus in the individual. The ancestral pencil converges upon the individual. The progenial pencil diverges from the individual. The practical consequences to the investigator of the fact that quite different possibilities are opened, according to which one of these two possible ways of studying genetic relationships one chooses, are extremely interesting, but time is lacking to go into their discussion in detail here. How profound in general they are is sufficiently indicated by comparing the achievements, in the way of advancing our knowledge of the hereditary process, of the biometric method on the one hand and the Mendelian on the other hand. It is

perhaps too much to say that the law of segregation and recombination would never have been discovered by study of the ancestry alone, but such a result would at any rate have been highly improbable.

The difference in viewpoint between the biometric and Mendelian methods has its most important practical consequence in the fact that it renders possible the application of the experimental method in the latter case, while practically excluding it in the former. One can experiment in regard to one's progeny but not in regard to one's ancestors. Herein lies the great advantage of Mendelian method as a mode of research. While Mendelism is philosophically a statistical method primarily, as has already been pointed out, it is not solely such, but instead formulates its problem in such a way as to permit the experimental mode of attack.

The essential thing which Mendelian studies of heredity do is to determine the *distribution* of hereditary differences amongst the progeny of a particular individual or pair of individuals. In other words, it studies the distribution of hereditary specificities. This is obviously a higher level of attack on the problem than the biometric. Because it is so is the reason that it has been so much more fruitful of results.

That the method is essentially statistical is evident. It deals with masses of individuals,

but they are offspring individuals, not ancestral. The generations which the Mendelist discusses are, in Bateson's now universally adopted terminology, the filial generations. From the standpoint of mathematics Mendelian statistical methods are crude and simple. Mendelism has unfortunately had no Karl Pearson to work out a special statistical technique directly adapted to the requirements of its data. Such a special technique may, however, be expected slowly to develop as time goes on.

I should like to digress here a moment to discuss a particular instance of the crudity of current Mendelian mathematics. In recent developments of Mendelian theory it has been a common practice to assume the existence of multiple factors [1] as the causal agents of a single character. There can be no *a priori* logical objection to this procedure. Its mathematical dangers are not usually perceived, however. Viewed as a logical method the multiple factor hypothesis is simply the well-known mathematical procedure of increasing the number of constants of a theoretical equation for the purpose of making a better "fit" to the data (in this case the observed ratios). But any one expertly acquainted with the general theory of

[1] This hypothesis was first used, at least on any considerable scale, by Nilsson-Ehle, "Kreuzungsuntersuchungen an Hafer und Weizen." *Lunds Univ. Arsskr.*, N. F. Afd. 2, Bd. 5, pp. 1–122, 1909. It has since been widely adopted by other Mendelian workers.

curve fitting knows that this process always runs a grave danger of becoming perfectly futile. Because if one increases the number of constants until it is equal to the number of classes of objects (ordinates) to be fitted, the "fit" is bound to be perfect, but also meaningless, because in no sense a true graduation. This danger is particularly grave in Mendelian work because there the number of classes or categories in the observed ratios is usually small, so that constants and ordinates may very easily become of the same order of magnitude. In this event agreement between observation and theory, however close, affords no critical evidence whatever as to the qualitative validity of the theory. (Cf. *infra*, p. 29.)

It has been shown by Yule,[1] Pearson,[2] and his students, and Hatai[3] that statistically the consequences of the distribution of hereditary specificities in accordance with Mendel's law are in essential agreement with the statistical results of the law of ancestral heredity. This result is at the present time chiefly of interest in that it furnishes the complete proof that hereditary differences are distributed in accordance with

[1] Yule, G. U. "New Phytologist," 1902.

[2] Cf. particularly Pearson, K., "Math. Cont. etc. XII," "On a Generalised Theory of Alternative Inheritance, with Special Reference to Mendel's Laws." *Phil. Trans.*, A. 203, pp. 53–86, 1904.

[3] Hatai, S. "The Mendelian Ratio and Blended Inheritance." *Amer. Nat.*, Vol. XLV, pp. 99–106, 1911.

those fundamental statistical laws to which other natural phenomena in general conform.

The most important achievement of the Mendelian method has been, I think, the demonstration that, on the one hand, hereditary differences behave, in the main, as discrete units, which are shuffled about and redistributed to individuals in the course of the hereditary process, to a considerable degree independently of each other; and on the other hand, that in typical cases this redistribution follows the simplest of statistical laws of dispersal, the point binomial.

Mendelism finds its limitations, just as did the biometric method, in the fact that from the logical standpoint it is essentially a statistical method which studies only the laws of distribution of things given or assumed. It examines only the distribution of hereditary specificities, and not at all, directly, their origin or determination. The former aim cannot be the goal of genetic science. A method which can travel only so far cannot hope to say the last word in the discussion of the problem of heredity. As a mode of research the Mendelian method of analyzing the progeny distributions rather than the ancestral will always be used. It was indeed one of the most brilliant methodological discoveries in the history of science. But it has limitations in the direction of what it can accomplish *per se* in elucidating the problem of heredity. Already Mendelian workers

perceive this fact. Thus Spillman [1] after pointing out that the facts of Mendelism can no longer be justly, or even decently, disputed, goes on to say (*loc. cit.*, p. 765): "The real trouble is not with the facts. It is with the interpretation of these facts. Just at present we have more facts of a certain kind than we know what to do with. We need some one to put meaning into these facts. We are in the position of a man lost in the wilderness. What he needs to find is a road. It does not make so much difference where this road shall lead, for all roads lead to each other. If he can find any road, it will lead him to where he can find people, and these can point out other roads leading more nearly in the direction he wants to go.

"In genetic investigations we need theories that will suggest lines of investigation that will be fruitful of results — that will lead, not to more facts of the kind we already have, but to new kinds of facts that will throw light on the subject from a new angle."

3. *The Cytological.*

In the field of genetics cytology is practically concerned with a single phase of the problem of heredity, namely gametogenesis (*B* of our initial analysis) and, to a somewhat smaller degree, with the initial stages of somatogenesis and the fertiliza-

[1] Spillman, W. J. "The Present Status of the Genetic Problem." *Science*, N. S., Vol. XXXV, pp. 757–767, 1911.

tion of the egg. The broad general result of the investigation of this problem is that certain definite nuclear components, the chromosomes, are distributed during gamete formation in a manner which parallels the distribution of hereditary characters as observed in Mendelian segregation. In other words, the chromosomes behave in gametogenesis as any structures which were the bearers of the causative agents of the inherited characters would be expected *a priori* to behave. This discovery is clearly one of first-class importance. It is justly to be regarded as one of the greatest achievements of modern biology. It furnishes strong grounds in favor of the basic conclusion that the determination of hereditary specificity is resident in the chromosomes. The familiar and widely accepted doctrine that the chromosomes are the exclusive "bearers" of hereditary qualities is a crude form of this conclusion. Those who have defended this doctrine, however, have been compelled from time to time to qualify their statement of it, because of the discovery of facts which either were absolutely difficult to reconcile with it, or were relatively more easily to be accounted for on some other hypothesis.

The greatest methodological difficulty in the case lies in the fact that cytology is essentially an observational and not an experimental science, though some brilliant beginnings in the latter direction have been made, notably by Boveri.

The cytologist is compelled by force of circumstances to use in the main static material — the dead stained cell. But in the problem of heredity he is concerned with dynamic phenomena — gamete and soma *formation*. It is obvious that the chances for error in interpretation are very great, and difficult to eliminate or control, when one is compelled to reason from a static, fixed structure, to a dynamic course of events.

It may fairly be said that a direct connection between a particular chromosome in the gamete and a particular structure in the adult has never yet been completely demonstrated. The closest approach to it, on the directly cytological side, is found in the case of the sex chromosomes, and here a causal nexus is not absolutely proven, though it seems doubtful whether by cytological methods alone it will ever be possible to get essentially nearer to a proof than we now are. The clear-cut and thorough researches of Wilson, in particular, and American cytologists in general, on the sex chromosomes appear to take the problem as far as purely observational methods can take it.

There has been no lack of cytological hypotheses regarding genetic phenomena in recent years. In particular it has become the practically universal custom to look for the explanation of divergent Mendelian ratios in cytological disturbances or deviations of some sort or other. This has conspicuously been the case in regard to the so-called

"linkage" phenomena of inheritance. The divergent Mendelian ratios obtained in such cases have been variously explained. Bateson and Punnett [1] have advanced the theory of reduplication of germ cells. This idea was first suggested in the following terms:

"The increase in number of the two types of cell, *AB* and *ab*, may be reached by proliferation of the two primordial cells of those two types. It may further be remarked that though the numbers characteristic of coupled systems cannot be produced by simple dichotomies, they can readily be represented as produced by a series of periclinal and anticlinal divisions. For example, if $AB^1$ by periclinal divisions give off $AB^2$, and this by anticlinal division become two cells, which again divide periclinally and anticlinally, seven cells *AB* are formed; by repetition of the same processes 15 are formed, and so on."

The direct cytological evidence in favor of the reduplication hypotheses would seem to be extremely meager. The theory can, to be sure, point to cytological observations which *might* furnish a basis for the genetic results observed. But this is a very different thing than a demonstration that they do furnish such a basis.

[1] The first statement of the reduplication hypothesis is found in Bateson, W., and Punnett, R. C., "On the Interrelation of Genetic Factors," *Proc. Roy. Soc.*, B., Vol. 84, pp. 3–8, 1911. Further developments of the theory have been made by these authors, Trow, Bailey, and others in recent papers in the *Journal of Genetics*.

One cannot escape the feeling that the statistical ratios themselves, derived from experiment and capable of direct verification, are the only really definite and substantial things in the case. The reduplication theory, in spite of the fact that it is stated in cytological terms, appears logically to be purely a statistical hypothesis for the graduation of a particular kind of observational data. Logically it falls in the same category of operations as fitting a curve to a series of physical observations. To this no objection can be raised. The only point is that any one who has had experience in the mathematics and logic of graduating observational data will not be inclined to rate the evidence in support of the qualitative *cytological* aspects of the hypothesis at quite so high a value as might seem at first thought to be warranted.

The logic of the case is possibly worth going into a little farther, since it involves what seems to me to be one of the most insidious and prevalent fallacies in biological research.[1] An investigator discovers some new and curious facts, which are capable of numerical expression. He then frames an hypothesis to account for them. Depending

[1] One wonders how prevalent the same fallacy may be in other sciences. The writer does not have an intimate enough familiarity with the actual facts to form anything more than a suspicion. The suspicion is, however, that it will not become either the physicist or the chemist to assume airs of superiority over the biologist on this point.

on the nature of the data and, perhaps even more on the nature of the person handling them, this hypothesis may take a directly mathematical form, as say the equation of a curve, or it may be seemingly quite unrelated to anything mathematical — as for example chemical, cytological, psychological, or what not. What the hypothesis is does not matter, *except in this respect that it always somewhere involves a statement or implication as to the qualitative cause of the quantitative phenomena observed.* The next step in the investigation is to calculate out for each particular observed case the *numerical* results to be expected on the basis of the hypothesis. These "expectations" are compared with the observations. If the agreement is good, the investigator is likely, and here lies the fallacy, to draw the conclusion that this agreement proves that the *qualitative* assumptions made in the hypothesis are correct. Of course the agreement logically proves nothing of the sort. The reason why it does not is found in the lack of uniqueness in the quantitative relations of qualitatively distinct natural phenomena.[1] Because two series of events follow the same curve it by no means follows that they are due to the same cause. A reasonable accordance of the numerical

[1] It should be pointed out here that in the present development of this argument I am drawing freely from a previous paper (*Amer. Nat.*, Vol. XLIII, pp. 302–315, 1909), in which the same point was discussed in relation to investigations on growth.

consequences of a particular hypothesis with the observed data does not prove, in the absence of another *kind* of evidence, that the hypothesis expresses the causal basis of the phenomena. If an hypothesis is true its numerical consequences must accord with observation: but the converse proposition that because there is agreement the hypothesis must be true, does not necessarily follow. It is one of the weaknesses of the human mind to fall into the error of thinking that it does: it is a mistake most of us have made in one form or another.[1]

Bateson's reduplication hypothesis seems to furnish an excellent concrete illustration of the point. Essentially the only evidence in favor of the hypothesis is that derived from the agreement between observed and expected statistical ratios. This is totally inadequate to base any cytological hypothesis upon. Some other *kind* of evidence must be forthcoming before it can be demonstrated that some gametes "reduplicate" to a just-sufficient degree to meet the exigencies of the case. An ingenious mathematician could probably frame

[1] One learns to be cautious about "expectations." There once came to my attention some results put together by a non-mathematical biologist, who had elaborated a very complicated mathematical hypothesis to account for his observations. The agreement was wonderfully close between "observed" and "expected." Some of its significance disappeared, however, when it was found, upon analysis of the hypothesis, that the mathematical methods involved were such that, barring an arithmetical error, *there could by no possibility ever be more than a fractional discrepancy between observation and calculation, whatever the nature of the observations!*

a dozen other hypotheses which would "fit" these aberrant ratios equally well. Furthermore, if he were also a biologist he could no doubt give each hypothesis a cytological flavor, and bring forward observations upon the chromosomes or other cell constituents in its support.

It is important in this same connection to consider carefully the results of Morgan regarding linkage phenomena. He and his students [1] have collected a great body of data regarding "linked" inheritance in *Drosophila*, which they interpret on an hypothesis which locates the causal determining factors for particular characters not merely in particular chromosomes, but further in particular places in the chromosomes. The ingenuity and keenness of insight displayed in the experimental analysis by which these results have been obtained must compel the unreserved admiration of any one who carefully examines them. This *Drosophila* work will stand as a masterpiece of experimental research. Logically considered the evidence in favor of the chromosomal hypothesis adopted by Morgan is based primarily on the general fact that the chromosomes, if they be assumed to behave in certain postulated ways, would furnish a mechanism which would account for the observed breeding results. The primary

[1] In a series of papers appearing during the last few years in the *Jour. of Expt. Zool., Science, Amer. Nat., Zeitsch. ind. Abst.-Vererb. Lehre,* etc.

objective data are the statistical results of Mendelian experiments. But, and herein seems to me to lie the logical as well as the factual superiority of Morgan's view over that of Bateson and Punnett, pains have been taken at every step to check the hypothesis by direct cytological observation as well as experimental evidence. The net result is that, up to the present time, everything that is known about the cytology of the germ cells in *Drosophila* is in entire and complete accord with Morgan's interpretation of the experimental results. The manifoldness and complexity of these experimental results have now become so great as to make it extremely improbable that this agreement between cytological and experimental data is fortuitous. On the contrary, the evidence comes extremely close to a logically complete experimental demonstration of two points; namely, (1) that the factors which have a differential effect in the determination of inherited characters are contained in the chromosomes, and (2) that the factors have a linear arrangement in these bodies.

The extraordinary difficulty of adducing anything like complete proof for the location of hereditary determiners in morphologically definite elements of the cell organization must be obvious to any one. The evidence is bound to be largely of an inferential character. On that basis the weight of evidence now available strongly indi-

cates that the chromosomes are directly and primarily concerned in the distribution and transmission of hereditary differences. This does not mean, however, that one can unreservedly assert that the chromosomes are the exclusive agents in hereditary determination. On the contrary, there are a number of facts recorded in the literature of cytology which, in the light of present knowledge, can only with great difficulty, if at all, be reconciled with such a view. In this connection it should be pointed out that the most acute and philosophical students in this field no longer regard the determination of hereditary specificities as probably confined *exclusively* to a single morphological element of the cell. Thus Wilson[1] has recently said: "In any case, the conclusion is not to be escaped, I think, that the whole cell system is directly or indirectly involved in the production of every hereditary trait. To treat the chromosomes as if they were central governing or controlling factors in the cell is a procedure of more than doubtful expediency. For the present, at least, all the requirements of investigation are sufficiently met if we think of the chromosomes, or that which they carry, only as *differential factors* in heredity, not as its primary or exclusive 'determiners.' Whether they possess a significance more fundamental than this is a question that may

[1] Wilson, E. B. "Heredity and Microscopical Research." *Science*, N. S., Vol. 37, pp. 814–826, 1913.

well await the results of further inquiry." Minot[1] expressed essentially the same conclusion in one of his Jena lectures. Doncaster[2] after reviewing the matter says that the evidence, "while not proving that the chromosomes are directly concerned in the transmission of inherited characters, makes such a hypothesis very plausible. Much unnecessary confusion, however, has arisen, from stating the hypothesis in the form — 'the chromosomes are probably the bearers of inherited characters.' Evidence has been adduced that the cytoplasm plays some part in determining these characters, and it has therefore been maintained that the statement is disproved. No one, however, would suppose that the chromosomes could act alone; they must act in and by their relation with the cytoplasm, and if the cytoplasm is that of a different species, the total effect must necessarily be different."

To summarize this section of the paper, it may be said that while cytology attacks directly one of the basic elements of the problems of heredity, gametogenesis, it finds immediate and serious limitations in two facts. The first is that the method of research in cytology is the morphologi-

[1] Minot, Charles S. "Modern Problems of Biology," Philadelphia, pp. 1–123, 1913.

[2] Doncaster, L. "Chromosomes, Heredity and Sex: A Review of the Present State of the Evidence with Regard to the Material Basis of Hereditary Transmission and Sex-Determination." *Q. J. Micr. Sci.*, Vol. 59, pp. 487–521, 1914.

cal, observational method. It is essentially a static method. It labors constantly under the very serious danger of error which inheres in inferring the course and nature of dynamic events by observation only of their static antecedents or consequences. In the second place, the technical difficulties of the material make it, in most instances, practically impossible to apply the experimental method directly to the cytological side of the problem of heredity. In spite of these limitations, cytology has made and will continue to make fundamental contributions to the progress of research in heredity. It is one of the essential methods of investigation in this field.

4. *The Embryological.*

Embryology has chiefly been cultivated for its own ends, which are, on the one hand, those of descriptive morphology and, on the other hand, those of a special field of physiology, *Entwicklungsmechanik.* Only in a relatively small proportion of instances has it been directly and purposefully used as a mode of research in genetics. Yet embryology is the science of somatogenesis, which was shown at the beginning to be one of the fundamental elements of the problem of heredity. It is a little difficult to understand why, with such splendid opportunities as the embryological method offers, so little light regarding the hereditary process seems to have come from the embryologist. To say this is not in the least

to depreciate the value of what has been gained. The studies of the last twenty years, particularly in the fields of experimental embryology and cell lineage, have accumulated a splendid mass of data regarding the principles of somatogenesis. From the standpoint of genetics, however, this material is in a scattered and diffuse state. What it needs is a synthetic treatment, in which all of the now scattered facts of experimental embryology shall be brought together and their relations to the general problem of heredity clearly shown. A few embryologists have, to be sure, contributed notably in this direction, chiefly in this country Brooks, and later Conklin, who seems to me in all his work to have perceived more clearly than any other recent student in this field that embryology has a very important, if not indeed its chief philosophical significance as a mode of attacking the fundamental problem of heredity. In this same connection the recent work of Gurwitsch[1] takes a high place.

There are several reasons why the embryological method has not in fact been more fruitful of generalizations of value in genetics. The first, and most serious, is the infrequence with which the working embryologist has had any real or deep appreciation of the relation of the problem

[1] Cf. for example this author's recent paper "Der Vererbungsmechanismus der Form." *Arch. f. Ent. Mech.*, Bd. 39, pp. 516–578, 1914.

of somatogenesis to the problem of heredity. He has been interested in other things. The second is that in the main he has worked with material of whose genetic history nothing was known. His material was "unpedigreed" material. The third is that embryology has generally studied only one part, the beginning, of somatogenesis. Post-embryonic development has been regarded as unworthy of consideration. Yet in the main it is with *adult* characters that the geneticist deals.

These seem to be the more important practical reasons why embryology has not hitherto been conspicuously illuminating as a mode of genetic research. Logically considered this method finds its chief limitation, like each of the other three methods, in that it takes hereditary specificities as things given, and makes no attempt, because it has no means of accomplishing such an end, to investigate their origin or determination.

From the methodological standpoint a sharp distinction must, of course, be made between the older purely descriptive embryology and the modern experimental embryology. It is the latter which is of the greatest potential value as a mode of research in genetics. The former (descriptive embryology) is subject to a considerable degree to the same practical limitation as cytology, namely it is essentially a static method. It is clear, however, that in experimental embryology, using the term in the widest sense to include the

study of such phenomena as regeneration, regulation, growth, etc., we have a first-class tool for the analysis of the principles of somatogenesis.

It is unquestionable that a splendid opportunity lies open for research in genetics by the embryological method, using the term in the broad sense just mentioned. It is altogether usual in current discussions of variation and heredity to neglect completely everything which comes between the two end terms of the ontogenetic series, the germ cell on the one hand and the adult soma on the other hand. But clearly what goes between is a most essential part of heredity itself. It is astonishing how little has been done on these extremely obvious problems. Dr. Boring and the writer[1] have recently undertaken a study of the regeneration of certain organs in heterozygous animals, to see what is the behavior of the successive regeneration in respect of the character for which the individual is heterozygous. Some of the results, as yet unpublished, are very interesting. For example, to mention but a single case, we find that an individual, heterozygous for plumage pattern, may from the *same* feather follicle at one time produce a feather having a pattern like that of one of the parents, and at another time a feather with a pattern like that characteristic of

[1] Cf. for a preliminary report of some of these experiments, Pearl, R. and Boring, A. M. "Some Physiological Observations regarding Plumage Patterns." *Science*, N. S., Vol. 39, pp. 143–144, 1914.

the other parent. Studies of this sort are bound to throw light on the puzzling — and too much neglected — problem of dominance.

## III. Conclusion

In what has preceded we have attempted a critical analysis of the four general methods which have been employed in the investigation of the problem of heredity. Two of these methods have been seen to be essentially statistical, and two essentially biological. The statistical methods — the biometric and the Mendelian — differ fundamentally only in that the former investigates primarily the ancestry and the latter primarily the progeny. Logically exactly the same distinction was found between the two purely biological methods — the cytological and the embryological. The former studies the ancestry of the germ cell (gametogenesis), the latter the progeny of the germ cell (somatogenesis).

All of these methods are valuable, and each has contributed to our present knowledge of heredity. No one of the methods alone can, however, solve the problem. They all have at least one fundamental limitation in common. This is that they offer no means of *directly* getting at any definite information regarding the origin, cause, or real nature of that specificity of living material which is the very foundation of the phe-

nomenon of heredity. The distribution of hereditary specificities, their putative morphological "bearers," and many other things about them have been studied more or less exhaustively. The things themselves have been speculated about, but not investigated to any but the slightest extent.

It is, however, plain, I think, what must be the mode of attack on this outstanding problem of genetics. In the present state of knowledge it is beyond dispute that the basis of the specificity of living substances lies in its chemistry. This plain fact has long been recognized by biologists, but until very recently this recognition has been equivalent practically to an *Ignorabimus*. The obvious complexity of the chemical processes going on in living material has made any direct investigation of the problem from this side seem hopeless to the biologist. But this period of despairing wonderment is passing, and that rapidly. The remarkable development of biochemistry in the last twenty years has put at the disposal of the geneticist a new technical equipment with which he may directly attack problems which formerly seemed impossible of approach. To mention but a single instance by way of illustration, the science of serology, which has so far been developed almost exclusively as an adjunct to medicine, puts into the hands of the biologist a technique of enormous possibilities in solving the most fundamental problem of genetics.

We have in the past not lacked chemical theories of heredity. One has but to mention the names of Pflüger, Verworn, Hatschek, Kossel, Adami, Guyer, and many others to recall such theories to mind. The difficulty, it would seem, under which all such theories have labored is that, in the absence of the developed biochemical technique now available, these theories have not been framed on a practical basis; that is, they have lacked the very essential property of being capable of direct experimental test. The time seems now ripe for a biochemical theory of the hereditary process, which in the light of our present knowledge and technical equipment in the fields of statistical mathematics, experimental breeding, cytology, and biochemistry, shall be adapted to experimental verification or disproof as the case may be.

Finally, if, as I fear may be the case, I have wearied the reader unduly with this discussion of methods, instead of conducting him on a journey into the more exciting field of results, my *apologia* must be that, however irksome and disagreeable the task, an occasional examination and overhauling of one's equipment is as essential to success in scientific operations as it is in military. The geneticist's working equipment is a good one, and has wrought well. I hope it has done no harm to try to see just what the limitations are to the usefulness of each tool in the list.

# CHAPTER II

## BIOMETRIC IDEAS AND METHODS IN BIOLOGY THEIR SIGNIFICANCE AND LIMITATIONS[1]

### I

THE last twenty years have witnessed the origin and development of what amounts to a new branch of biological inquiry; namely, biometry. This subdivision of biological science, which has within this period come to be practically a distinct and separate *Fach*, may fairly be said to have taken its origin at about the year 1895 in the pioneer investigations of Pearson and Weldon. In making this statement there is no implication that there had not been important quantitative work in biology, of one kind or another, before 1895. There certainly had been a considerable amount of such work. It had, however, fallen in special and rather restricted fields. Most important in this earlier quantitative biology are probably to be reckoned the studies of the anthropologists. In this field the work of Quetelet and of Galton stands preëminent. The work of these men, and in particular that of Galton, indeed served in

[1] This paper, in somewhat modified form, was first published in *Scientia*, Vol. X, 1911.

considerable degree as the stimulus for an extension of quantitative ideas and methods into other and broader fields of biology.

Nor was physical anthropology the only phase of biology which had been definitely cultivated along quantitative lines before 1895. Certain branches of physiology have long been highly developed in this direction. One thinks particularly in this connection of the study of the physiology of nerve and muscle. It is probably no exaggeration to say that so complete and detailed a knowledge of the quantitative as well as qualitative relations of the phenomena involved as we possess for nerve-muscle physiology does not exist to-day in any other branch of biology.

But both in anthropology and physiology quantitative methods of work were only an incident. In many cases, as we can now see, the mathematical methods used were inadequate, or incorrect. Until Pearson, following the pioneer investigations of Galton, began his work no one had ever made any consistent and systematic attempt to develop a special calculus directly adapted to the discussion and analysis of biological data. Because he did just this thing Karl Pearson will stand as the founder of biometry as a definite technical subdivision of the science of biology.

The publication of the earlier memoirs in Pearson's fundamental series of *Mathematical Contri-*

*butions to the Theory of Evolution* in the period from 1895 to 1900 excited much interest among biologists all over the world. The possibilities opened out by these new methods of research were quickly noted, and served as a stimulus to start a number of workers off on the new paths. The growth of interest and the accumulation of results on the subject were such as to necessitate the launching of a new journal, *Biometrika* (now in its tenth volume), devoted entirely to the publication of researches in this field. Thus, biometry started on its career as an independent twig of the biological tree with a very vigorous and apparently healthy initial growth.

Events proved, however, that much of this growth was not by any means of a healthy character. A very decided reactionary feeling on the part of biologists regarding biometry began to make itself felt. In place of the enthusiasm with which efforts toward a quantitative biology were at first received, an intensely critical and in some cases distinctly hostile attitude appeared. Students were advised to have nothing whatever to do with this iniquitous biometry. It was felt that biometrical literature could safely be, and to a large extent was, disregarded by the general biologist.

The reasons for the development of this reactionary attitude are now, I think, tolerably clear. The first and most important was that biologists

very generally misunderstood the significance and possibilities of biometric methods as applied to their subject. At the outstart the enthusiasm of the biometric workers led to great expectations as to what the new science was going to accomplish. Because these expectations were in large degree based on an entire misconception of what mathematical methods could by any possibility do, they were not fulfilled, and this naturally led to more or less of a feeling of aversion to the whole subject. Such a result would have been inevitable whatever the quality of the biometric research done.

In the second place, biometry was, for some time, taken to be a school of biological philosophy rather than what it really is, a method of research. The great activity of biologists during recent years in the analytical study of inheritance by the method of experimental breeding has served to establish on a firm basis certain fundamental principles of the physiology of the hereditary process (the principles of segregation, and of the normal stability of homozygous strains). It is further the fact that certain views regarding the method of evolution and of inheritance in plants and animals which have been upheld by certain leading biometrical authorities are, in regard to some fundamental points, utterly at variance with the results of these experimental investigations. By a regrettable confusion of thought,

biometry has been arraigned for the views on purely biological topics held by certain individuals. Surely the application of appropriate mathematical methods to aid in the solution of biological problems involves *per se* the maintenance of no particular theoretical position whatever with regard to the fundamental nature of biological processes or phenomena.

In the third place, it must be admitted that a good deal of the early work in biometry was of a superficial character, and made no contribution of moment either to biology or to biometric theory or practice. On account of the novelty of the view point an inexhaustible wealth of material lay ready to the hand of the biometrician. If a knowledge of the statistical facts of variation in organism *A* was a good thing to have, a similar knowledge for *B* ought also to be a good thing. Nothing could be simpler than to collect a lot of individuals, measure or count something, and then write a paper setting forth the results of this activity. Too often the enthusiasts who rushed into the new line of work seemed to expect the application of quantitative methods not only to solve problems, but also to furnish problems ready made.

Biometry has now safely passed that stage in development, which every new line of investigation passes through sooner or later, in which it suffers at the hands of its overzealous friends. It

is gradually coming to be clearly recognized by general biologists that biometric methods when properly used add an important weapon of research to the investigator's armament. It is the purpose of this paper to attempt in a very modest way to help along, if possible, this better understanding of and greater sympathy towards biometric work.

## II

The underlying and essential point of view of biometry has been quite generally misunderstood by biologists. In the first place biometry is often strongly and quite unjustly criticized because it has developed primarily as a statistical science. It is supposed that this method of inquiry cannot properly or profitably have anything to do with any problems not immediately reducible to frequency polygons and correlation tables. The charge is made that biometrical methods can deal only with mass phenomena, and that they intentionally disregard the detailed study of the individual, and therefore lead directly to experimental indeterminism as a mode of biological thought. Such a charge is based on an entire misconception of the biometrical standpoint. This attitude, however, has done a great deal of harm in hindering the wider use of these methods by biologists. Nothing has been more strongly emphasized by the trend of recent biological discovery than the

importance of the thorough, searching study of the individual, whether the matter under investigation be behavior, inheritance, variation, or morphogenesis. It is no recommendation for a new line of scientific inquiry to be supposed to ignore or belittle this fact.

A second misconception which prevails rather widely is to the effect that biometric methods, being supposedly entirely statistical in character, necessarily require as a working postulate at the outstart that the accumulation by selection of small fluctuating variations is the primary and fundamental, if not the sole method of organic evolution. Such an idea is, of course, absurd. The purpose of biometrical study so far as it is applied to the evolution problem is precisely to find out what has been the method of evolution. Whatever the results of such inquiry may turn out to be has no relation to either the validity or usefulness of biometric methods *per se*. The statistical methods or calculus developed by Pearson are particularly adapted to the study of variation of the continuous fluctuating type, but one who uses this calculus is in no way compelled because of this fact to take any particular position in regard to the theoretical question of the biological importance or significance of this kind of variation in evolution. This is a matter to be settled by direct experiment and observation. If, as many biologists are coming to believe, this type of

variation has, in a strict sense, very little if any significance in evolution, biometric methods will help to demonstrate the fact. Further, as has been implied above, the statistical study of variation is only one side of biometry.

Based on a misconception similar to that just discussed is the point of view which criticizes biometry as being necessarily tied fast to a particular view regarding the hereditary process. The "law of ancestral inheritance" first enunciated by Galton and later extensively developed by Pearson is simply a statistical statement. It concerns itself with the end results of the action in a general population of a whole complex of biological processes, of which inheritance is only one. It is the opinion of many workers in the field of genetics that this "law" probably has very little direct relation to the really significant biological facts of heredity, and that whatever apparent significance it may have is largely accidental and fortuitous. But whether this opinion is correct or not certainly has no bearing on the question of the validity of bringing appropriate and correct mathematical methods to the aid of the investigator wherever they can be of help in solving problems. It is a confusion of thought to criticize a scientific method of investigation because of the theoretical views held by some of those who employ it. There is yet to be discovered a scientific method which can be

depended upon to give correct results invariably, regardless of how it is handled or applied. Biometry is no worse off in this regard than is the experimental method, the position of which in biological methodology is now so secure.

The real purpose of biometry is the general *quantification* of biology. Its fundamental viewpoint is that without a study of the quantitative relations of biological phenomena in the widest sense it will never be possible to arrive at a full and adequate knowledge of those phenomena. This viewpoint insists that a description which says nothing about the magnitude of the thing described is not complete but, on the contrary, lacks an element of primary importance. It insists, also, that an experiment which takes no account of the "probable error" of the results reached is inadequate and as likely as not to lead to incorrect conclusions. Further, and more broadly, it is certain that not only are quantitative methods needed in biology, but also that a far more serious need is for something of the methodological viewpoint — the mode of thinking — which is characteristic of the exact sciences. What the writer conceives to be the true and basic standpoint of biometry cannot be better expressed than in the following remarks of that master of an exact physical science, Lord Kelvin, in an address on *Electrical Units of Measurement.*[1] "I often say

[1] "Popular Lectures and Addresses," Vol. I, p. 73.

that when you can measure what you are speaking about and express it in numbers, you know something about it, but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts, advanced to the stage of *science*, whatever the matter may be."

Having taken this position, the next thing in order obviously is to develop mathematical methods especially adapted to the treatment of biological data. Such a step is no more to be criticized than is the demand of the experimentalist that he shall have apparatus adapted to his needs, or of the morphologist that he shall have the latest and best type of microscope for his most detailed and important researches. It is the most obvious right of an investigator that he shall have highly developed and adapted technical aids whatever his field of work. In accord with this principle there has been developed (partly by borrowing from pure mathematics and partly *de novo*) a very efficient and tolerably complete system of special mathematical methods particularly adapted to the analysis of quantitative biological data.

Unfortunately the more recondite of these methods cannot be understood at all by the general biologist unless a considerable amount of careful and thorough study is given to them.

Even the simpler of current biometric methods are not fully understood by the majority of biologists, nor can they except through special study of their mathematical origin and development. But is there any reason why the biologist should expect to have intuitive comprehension of these methods? No one would expect to apply successfully the complicated and delicate surgical technique of Pawlow or Carrel to the solution of biological problems without careful preliminary study and practice of these methods, continued till they were really mastered. The case is not different with any other higher development of scientific technique.

Because of the lack of a full comprehension of the meaning and significance of the mathematico-statistical methods used in biometry, these methods have been subjected to a great deal of unreasonable and futile criticism. It is argued that these methods are in large part worthless because they are too refined. Biological data are held to be of so coarse and inaccurate a character as to make any but the roughest kind of treatment of them of no significance. Such a view misses entirely the purpose and meaning of the biometrical calculus. It is just because biological data necessarily *are* often rough that we need refined mathematical methods in their treatment in order to test and check the conclusions to be drawn from them, and in order to show their

true trend and significance. An example will help to make the point here clear. Mortality statistics are usually available only in units of years of life. This is a rough unit. For actuarial purposes it is desirable to know, for example, the probable duration of life much more accurately than in terms of years. It is possible to get this information, accurate to a very high degree, by the application of appropriate mathematical treatment to the rough yearly data. In this connection, too, is to be considered the frequently made statement that no statistical constant can be more accurate than the data on which it is based. It is very easy to demonstrate that this is not true.[1]

While, as has been pointed out in this section, the methods and point of view of biometry have not always been understood, yet the indications are that matters are improving very rapidly in this respect. In particular the great interest and activity now being displayed in the study of inheritance and plant and animal breeding is doing much to increase the use of biometric methods. Breeders are accumulating masses of data which they wish to analyze. To do this necessitates in many cases the use of biometric methods. As the really purposeful employment of these methods to help solve practical problems increases they must inevitably come to be better understood by the great mass of biologists.

[1] Cf. Pearl, R. *Amer. Nat.*, Vol. XLIII, pp. 238–240, 1909.

## III

Let us now turn to a consideration of some of the ways in which biometrical methods may be of immediate value to the progress of biology. What has biometry to offer to biology that is useful? Or, in other words, what is the significance of biometry?

Stated most broadly it may be said that biometry brings to biology a fairly well developed method or system for the more precise, accurate, and complete *description* of biological phenomena. Biometry is, in last analysis, a descriptive method. Like all other descriptive methods, or phases of science, it is not likely in and by itself ever to solve completely any problems. It must always work in conjunction with the experimental method to attain the highest (*i.e.*, most valuable) type of results. But at the same time it brings to the aid of the experimentalist that which is of the utmost importance; namely, an adequate method of describing, analyzing, and in general reaching correct conclusions from the results of experimentation.

Biometry affords an extension of descriptive methods in a direction where such an extension is often an absolute requisite of truly scientific work, and in a direction where no other method is available. Ordinary biological observation and description has as its unit the individual, or some part

of the individual. It describes the individual in terms of its own qualities or attributes. If an adequate description of the individual is given, it does not concern itself solely with the separate parts, but discusses the qualities and attributes of the individual as a whole. Thus it would not be an adequate description of a man to say that he was the sum of such and such bones, muscles, nerves, and blood vessels. A cat possesses much the same bones, muscles, nerves, and blood vessels that a man has. It is probably no exaggeration to say that the similarity between man and cats in respect to these organ systems is so great that a person without previous experience of either — say a Martian — unless he were deeply versed in anatomy, might very possibly consider cats to be dwarf men if he had as a basis of distinction *only* a formal description of the organ systems named. Any adequate description of an organism must include as its most fundamental and important part an account of the attributes and qualities of that organism as a whole.[1]

Now a little consideration will serve to convince one that the ordinary methods of description as used in biology fail (*i.e.*, become altogether inadequate) when the attempt is made to deal with any *group* of individuals, as for example a population,

[1] This point has been developed in a masterly way in a paper by Professor William E. Ritter, having the title "Life from the Biologist's Standpoint" in *Popular Science Monthly*, August, 1909.

race, variety, species, or larger group. These methods fail because they are fundamentally and necessarily incapable of giving a description of the group (whatever its magnitude) *in terms of anything but the individuals which compose it.* That is to say, they have no way of getting at a description of a group (*e.g.*, a species) as a whole or as such, in terms of its (the group's) attributes and qualities. Let an illustration make this point clear. The purpose of systematic zoölogy is to classify and arrange animals in natural groups. As a necessary step in the carrying out of this purpose it is obliged to attempt to define, which means describe, these groups. But its whole way of going about this process is a confession of the fundamental inadequacy of the method. The systematist frankly makes no attempt whatever to describe or define a particular species as a species (*i.e.*, as a group of animals) in terms of its (the species') qualities. Instead he describes one individual animal belonging to this species; affirms either expressly or tacitly that all other individuals belonging to the species are "about" or "generally" like the individual described, and then calls the net result the definition or description of the species. But now surely this is not a description of the *species* at all. An adequate description of the species will be one which takes account of its peculiarities as a unit, and indicates how it as a unit or as a whole

is distinguished from other similar groups. In making this statement there is, of course, no implication that the facts set forth by the systematist are not desirable and useful. But something more is needed to gain a well-rounded, adequate idea of the group, whether species, variety, or any other.

It is a particular and fundamental point of significance of biometry for biology that it offers an adequate solution of just this problem of the description of the group as a whole or a unit in terms, not of its component individuals, but of its own attributes and qualities. The biometrical constants (mean, standard deviation, coefficient of variation, etc.) are, within the limits of error of random sampling, constants characteristic of the group as such, and not of any particular individual or individuals in it. So, further, the shape of the variation curve for a particular group of organisms is something definitely characteristic for the group.

The fact that in statistical methods we have the means of accurately describing the attributes of groups or organisms as groups affords an opportunity of investigating why groups (*i.e.*, species, varieties, etc.) come to have the characteristics which they do. It is the highest aim of the biometrical study of variation to determine the biological causes which underlie the formation of the particular sort of frequency curves which actually are found, rather than some one of the innumerable other sorts which might conceivably

have arisen in any individual instance. A first study in this direction was made by the writer for variation in the plant *Ceratophyllum.*[1] More recently fundamental researches of a similar character have been made by Jennings on variation in *Paramecium.* In the writer's laboratory a study has lately been made, from this point of view, of the inter-individual (racial) variation curves of one of the higher plants.[2] The first necessity in all such analytical studies must be a precise description and definition of the things which are to be analyzed. Such a description the application of biometric methods furnishes.

In what has been said regarding biometry as a method of group description, reference has been made, for the sake of simplicity of illustration, to groups of individual organisms. The same considerations, however, apply with equal or perhaps ever greater force to the study by biometrical methods of groups of like parts or organs within the single individual. Appropriate quantitative methods make it possible to detect and analyze the most subtle phenomena of differentiation in the development and growth of the individual. In the absence of methods for dealing with a group of parts or individuals as such, one

[1] Carnegie Institution of Washington, Publication No. 58, 1907.

[2] Pearl, R. and Surface, F. M. "Growth and Variation in Maize." *Zeitsch. f. ind. Abst.- u. Vererbungslehre,* Bd. XIV, pp. 97–203, 1915.

would be quite unable to attack a whole series of interesting and fundamental problems of morphogenesis. Such problems, for example, as the precision of morphogenetic localization, or the degree of variability of successively regenerated structures (does the morphogenetic mechanism learn to work better with practice?), or the existence and nature of fundamental laws of growth determining the general features of the tectonics of large groups of plants and animals, etc., demand the application of biometric methods if they are to be adequately treated. There is a host of problems of morphogenesis of this general character as yet hardly touched at all by the biologists. Biometric methods which enable us to deal with groups of things or events as groups furnish the key to the successful attack on these problems. Pioneer work in this direction is being done, but there is a vast and fertile field here, the proper cultivation of which will demand the combined efforts of many workers. To the application of appropriate biometric methods in this field we may confidently look, I think, as the source of a significant advance in the building up of a science of causal morphology.

The second fundamental contribution of biometry to biology lies in the fact that it has shown, and in the future will still more impress upon the biologist, the significance and great importance in all his work of the probable error concept.

Whatever the material of scientific investigation, whether animate or inanimate, it is a fact of universal experience that just as soon as observation or experiment concerns itself with any quantitative aspect of a phenomenon it is impossible ever to get precisely the same result twice. The more refined and delicate the instrument, and the finer the units in which the measurement is made, the more evident does it become that the "absolute" determination of any magnitude whatsoever is humanly impossible. The problem of all quantitative science, therefore, is to determine with a maximum of accuracy the probability that any particular unknowable magnitude lies within any assigned limits. One never can say, and be scientifically accurate, that a particular stick is precisely 11.5 cm. long, but if it be worth one's while, it is possible to determine the mathematical probability that the true length of the stick lies between say 11.498 and 11.502 cm. In the writer's opinion it must be regarded as the point of greatest value of statistical theory for science in general that it furnishes the method of determining such probabilities.

That the probable error concept is of high importance for biology is so evident as not to need lengthy discussion. When one considers what a large part of the results of experimental investigations of all kinds of physiological topics (to take but one instance) are quantitative in charac-

ter, and based on the observation of relatively few individual cases, the significance of probable error determinations in that field is clear. Similarly, in the experimental study of inheritance along Mendelian lines the results, as has been pointed out in the preceding essay, are quantitative and statistical in character. A statement of such results without probable errors is incomplete.

In the third place biometry gives us a method of measuring the relationship between the phenomena, in the multitude of cases where this relationship is not of a simple cause and effect sort. In observed biological phenomena there very often exists a correlation rather than a strictly causal relation between events or characters. The reason for this doubtless lies in the fact that in biology we are dealing in most instances with complex phenomena. As the complex phenomena are analyzed into their simple (or, perhaps better, simpler) components, correlation passes over into causation. It is, however, a great aid in making such analyses to have a method of measuring exactly the degree of correlation which exists between observed phenomena or characters. Of such methods current biometrical technique furnishes a great wealth. It is possible now to determine the degree of correlation or association existing not only between physically measurable characters, but also between qualitative characters not capable of precise measurement.

## IV

Let us now turn our attention to the consideration of the limitations of biometry as a branch of biological science. There is, I think, fundamentally but one such limitation of a *general* character. This arises out of the fact, already mentioned, that biometrical methods of research are, in last analysis, strictly and purely descriptive in character. There are but two general ways of acquiring and formulating a knowledge of natural phenomena. These are the descriptive method on the one hand, and the experimental method on the other hand. Biometrical methods belong in the first of these categories. The only thing which they are able to do is to furnish a description, in quantitative terms, of existing phenomena. This does not, of course, imply that they are not useful aids in experimental investigations. Indeed, it is just here that, in the writer's opinion, biometry finds, in general, its highest usefulness in biology. It is only desired to emphasize the fact that biometric methods are *per se* purely descriptive, and have the limitations implied thereby.

In actual biometric work the importance of keeping clearly in mind the limitations and precise significance of the methods used is great. To reach biologically significant results one must understand clearly just what is being measured

and what the utility of that measurement really is. Failure to do this is bound to result in confusion of thought. No description has any significance unless the thing described has some meaning and interest. Just here is where so much of the biometrical work which has been done has failed. Too often has there been an espousal of the forlorn hope that the application of biometric methods might inject biological interest and meaning into a problem previously quite destitute of these attributes.

Further, it is of the highest importance for the correct application of biometric methods to understand thoroughly the biological implications of the particular method employed. Failure to do this is bound to lead one into all sorts of pitfalls. It has been a very unfortunate boast of some biometricians that their methods involved no biological assumption or implication whatever. Such a statement is seen upon critical examination to involve a logical difficulty. Biometric methods, considered solely as pure mathematical reasoning, certainly have no biological implications, but the moment they are applied to biological data for the solution of biological problems they do carry biological implications. Otherwise their application is altogether irrational and futile. If no biological meaning or implication attaches to the determination of the degree of correlation between parent and offspring, for example, it is

evidently a waste of time to calculate such correlations. Further, if it cannot be clearly shown that the method of determining such correlations is such as to lead to a biologically valid result, the application of the method in biology is equally idle. As the point here under consideration is one of fundamental importance, it will be well to give it full discussion with a concrete illustration.

It has been frequently maintained by Pearson [1] that the "law of ancestral inheritance," which states that the correlation between offspring and their ancestors decreases in a geometrical progression as the number of included ancestral generations increases, involves no biological implication whatever regarding inheritance. Further, it has been stated on the same authority that the method by which this law is deduced (namely, by determining the correlations which exist between offspring and their ancestors) is valid whatever may be the biological basis or mode of inheritance. Now, as a matter of fact, practically all of the work which has been done upon inheritance by Pearson and his co-workers seems to the writer to involve from its very beginning a fundamental biological assumption. This assumption is that a correct determination of the correlation in respect to external, *somatic* characters between genetically related individuals, is an adequate measure of the

[1] Cf. for example, *Biometrika*, Vol. II, p. 217, 1909.

intensity [1] of inheritance between these individuals. But the validity of this assumption has never been demonstrated, and presumably never can be, because the assumption itself is contrary to demonstrated biological facts, which can at any time be experimentally verified. The facts to which I allude are those upon which rest the demonstration of the existence of the *genotype* as contrasted with the *phœnotype* in inheritance.[2]

These facts in general show that the somatic and the germinal conditions or states with reference to a particular character may be quite different in the same individual. It results, then, that the somatic condition of such a character in the progeny has no direct or necessary relation to the somatic condition of the same character in the parent.

Nothing is brought out more clearly by all recent experimental studies of inheritance than that the somatic condition of a character in a particular organism is a very unreliable criterion

[1] This term "intensity of inheritance" has been very generally used by biometricians. One ventures to wonder, in the light of present knowledge, whether the expression itself does not involve an unsound assumption. As a matter of fact, is inheritance ever a *graded* phenomenon, as implied in this expression? The whole body of evidence available makes this seem to me extremely doubtful. However, I have no desire at this time to press, or even to argue, the point. I merely offer the suggestion that the critical reader think it over for himself.

[2] Cf. Johannsen, W. "Elemente der exakten Erblichkeitslehre," 3d edit., 1913.

of the probable condition of that character in the progeny of that organism.

Thus, to take some concrete instances by way of illustration, if one breeds a Cornish Game cock bird to a Barred Rock female, the female offspring resulting will not be, in respect to plumage, color pattern, in the slightest degree like either parent. On the contrary they will be solid black in color.[1] A knowledge of the somatic condition of the parents, in such a case, no matter how detailed it might be, would, in advance of the actual breeding test, give no clew whatever as to the probable somatic condition of its offspring. Yet, in its procedure of correlating parent and offspring in respect to somatic characters, the "law of ancestral inheritance" definitely assumes that the somatic condition of the one may be taken as a valid and reliable criterion of the other. Many cases like that just cited might be given from existing Mendelian literature. Indeed, it may be said that nearly all experiments in crossing plants or animals furnish illustrations of the general principle that the somatic condition of a character may be a most unreliable criterion of its probable behavior in inheritance.

It may be objected that the types of cases cited in the preceding paragraph are hardly a fair basis for the criticism of the law of ancestral inheritance

[1] For details see Pearl, R., and Surface, F. M. *Arch. f. Ent. Mech.*, Bd. 30, pp. 45–61, 1910. (Roux Festschrift.)

in regard to the logical foundation of the procedure by which it is deduced, since the illustrations mentioned concern themselves with qualitative variations, whereas it is with quantitative variations that the biometrical study of inheritance has had to do. As a matter of fact the case for the law is made no better if this contention be granted, though it is difficult to see what reason exists for supposing that so-called qualitative variations are not inherited in fundamentally the same way as are so-called quantitative variations. Every character and every variation has both a qualitative and a quantitative aspect. But if we consider only the quantitative aspect of the matter, as has been done in much of the biometric work on inheritance, the same principle of the unreliability of somatic conditions as a criterion of hereditary behavior comes clearly forth from the work of Johannsen[1] on beans, that of Jennings[2] on *Paramecium*, investigations regarding the inheritance of fecundity in the domestic fowl,[3] and many other recent studies along similar lines by various investigators.

All of the experimental investigations referred to agree in showing in a most definite and indubitable manner that there exist two distinct cate-

[1] *Loc. cit.*

[2] Jennings, H. S. *Proc. Amer. Phil. Soc.*, Vol. XLVII, pp. 393–546, 1908.

[3] Pearl, R. *Jour. Exp. Zoöl.*, Vol. 13, pp. 153–268, 1912.

gories of variation, and that this fact must never be lost sight of in any discussion of heredity which is to lead to valid conclusions. On the one hand are the variations which are definitely inherited (*i.e.*, reappear in the progeny), presumably because they are in some way represented in the germinal substance; on the other hand are the purely somatic variations which do not reappear in the offspring and are not inherited, presumably because they are not represented in the germinal substance. Now the "law of ancestral inheritance" entirely disregards the existence of these two sorts of variations. In its fundamental thesis that the correlation between parent and offspring in regard to somatic conditions is a valid measure of the intensity of inheritance it definitely and implicitly assumes that all variations are of equal significance in heredity. Upon this fundamental biological error, which is taken as a basic assumption, the whole superstructure of the biometric treatment of inheritance is reared. When the significance and consequences of this initial error are perceived it is seen at once that the whole reasoning, so far as it concerns heredity, falls to the ground. Thus it is assumed that the existence of a definite degree of correlation (say $r = 0.40 \pm 0.02$) between parent and offspring indicates inheritance, providing both generations have been reared under reasonably the same environmental conditions, and an absence of correlation

($r = 0$) under similar circumstances means that the character studied is not inherited. But the work of Johannsen and Jennings indicates that in general there is no correlation ($r = 0$) between parent and offspring within the genotype (*i.e.*, within the same pure line). Are we to conclude then that there is no *inheritance* within the genotype — within a group of individuals where the offspring breed absolutely true generation after generation?

What a correlation coefficient deduced from a parent-offspring correlation table which includes a random sample of parents in general, or offspring in general, really measures, is what may be called the orderly heterogeneity of the material which goes into the table. One gets out of the table by way of the correlation coefficient merely a measure of what was put into the table as raw data. If the table includes individuals belonging to several distinct genotypes, we may expect to get from it, in many cases at any rate, a sensible correlation coefficient between parent and offspring. But this coefficient does not measure the intensity of inheritance between parent and offspring with reference to the character considered. It simply measures the mutual interrelation (as to range and distribution of variation) of the several genotypes which went into the table. The coefficient indicates, in other words, that there are certain groups of individuals within the table which are

differentiated, in respect to both parent and offspring, from certain other groups in the same table. It does not tell us what the basis of this differentiation is. It may be inheritance, as it is assumed to be in the illustration here under discussion; it may be local environmental differences or it may be anything whatever so far as the correlation method *per se* helps us. The only way to determine whether the "differences" indicated by the correlation method are really heritable is to apply the method of individual pedigree analysis to the complex, heterogeneous material of the table. If it is possible to isolate and propagate distinct genotypes from the material, then it may be concluded that the primary basis of the differentiation or heterogeneity detected by the correlation coefficient was inheritance.

There can be, of course, no valid objection to the study, in and for itself, of the correlation existing between genetically related individuals in respect of somatic characters. Such studies may, indeed, for one reason or another, have a high intrinsic interest. The point I would emphasize, however, is that in dealing with such correlations one should always keep clearly in mind that he is not dealing directly and primarily with phenomena of inheritance, but only indirectly and secondarily.

To summarize this discussion, it may be said that to attempt to draw conclusions in regard to

inheritance from studies involving the correlation method alone is futile, because the coefficient of correlation in such studies can only tell us of the existence and degree of an orderly differentiation or heterogeneity in the material collected together in the correlation table. To determine definitely whether the basis of the heterogeneity is inheritance, recourse must be had to pedigree analysis in terms of individuals and of families. Of course it will be true in many cases in actual practice that general considerations make it *a priori* highly probable or even certain that the basis of an observed heterogeneity is inheritance. In such cases the coefficient of correlation may be said to indicate the existence of inheritance, though not to measure its intensity. But the first and fundamental demonstration that a character is inherited in a given group of individuals must always rest upon some other basis than coefficients of correlation alone. The essential point here made regarding the inherent difficulty in the interpretation of correlation coefficients, has been very ably discussed by G. Udny Yule.[1] I am in entire agreement with his position.

The general purpose of this discussion regarding the study of inheritance has been to show, by a concrete example, how erroneous biological assump-

[1] Yule, G. Udny. "On the Interpretation of Correlations between Indices or Ratios." *Jour. Roy. Stat. Soc.*, Vol. LXXIII, pp. 644–647, 1910.

tions at the foundation may invalidate completely a whole line of biometrical reasoning. It is hoped that this discussion will emphasize the importance of the point made at the beginning, that it is necessary always to keep clearly in mind the limitations of the biometric method in general. Biometry furnishes a valuable and refined extension of the descriptive method. Its future usefulness to biology depends largely upon a clear recognition of two things. These are, on the one hand, that biometric methods are, in their essential nature, purely descriptive methods of investigation, and, on the other hand, that for the valid employment of these methods, it is necessary to use as much general biological intelligence in regard to the significance of the biological problem, the biological validity of the assumptions made in applying mathematical methods to the problem, and the applicability of the particular methods used to the particular problem studied, as would be exercised in any investigation by any other method.

# CHAPTER III

## ON THE NATURE OF STATISTICAL KNOWLEDGE

### I

THERE is a very general tendency, including in its operation not only the layman but also the professional man of science, towards the notion that there is a special virtue, a sort of transcendent heuristic worth, in such knowledge as is reached by the examination of large numbers of cases. There seems to be a feeling, sometimes apparently almost mystic in its origin and in its strength, to the effect that statistical knowledge is a higher and better kind of knowledge than any other. Numberless quotations might be cited to show the prevalence of this view. Every one has seen passing, as it were in review, the line of problems, which, if we may trust the assertions of the interested individuals, can "only be solved" by the application of the statistical method. Evolution, and the factors of evolution, variation, heredity, and so on, are conspicuously the biological problems of which this assertion has been made.

Now this attitude towards statistical knowledge and statistical ideas (which, of course, include

besides the compilation of large numbers of individual instances, the concepts of averages, approximation, and probability) may be entirely right and justifiable. Indeed, a cautious person is bound to be very chary about even suggesting any criticism of it when he finds the following statement made by so distinguished an authority as Professor Royce.[1] "I am next led to say that whether the natural world is a mechanism or not, the statistical view of nature would be, and so far as we know the facts, is applicable to sufficiently complicated systems of things and events, not as mere substitute for those more exact computations which our ignorance of mechanical laws makes necessary, but as an expression of a very positive, although only probable and approximate knowledge whose type all of the organic and social sciences, as well as most aspects of the inorganic sciences, illustrate. There is, therefore, good reason to say that not the mechanical but the statistical form is the canonical form of scientific theory, and that if we knew the natural world millions of times more widely and minutely than we do, the mortality tables and the computations based upon a knowledge of averages, would express our scientific knowledge about individual events, much better than the nautical almanac would do. For our mechanical theories

[1] Royce, J. "The Mechanical, the Historical and the Statistical." *Science*, N. S., 1914.

are in their essence too exact for precise verification. They are verifiable only approximately. Hence, since they demand precise verification, we never know them to be literally true.

"But statistical theories, just because they are deliberate approximations, are often as verifiable as their own logical structure permits. They often can be known to be literally, although only approximately, true."

It is the object of the present paper to discuss statistical concepts and statistical methods, for the purpose of trying to see whether these methods do in fact seem to lead to a higher kind of knowledge, and are hence to be considered destined to become the exclusive methods of science. In this discussion let us endeavor to avoid dogmatic assertion, since, in the first place, assertion does not really get us far in the search for truth, and, in the second place, the writer himself feels in regard to this question very far from that serene consciousness of being quite unassailably right which is essential to proper dogmatism. Indeed, it is for the purpose of definitely formulating some doubts, which have grown in the writer's mind during fifteen years in which the greater part of his time and energy have been devoted to the application of statistical methods to biological problems, that this paper is written. Very likely some will not agree with its reasoning or its tentative conclusions, but even in such event, it

may help the disagreeing reader to the more complete classification of his own ideas about statistical concepts.

## II

Let us first consider this question: What caused the development of the statistical viewpoint and method, which in science had such an important growth in the nineteenth century? For what purposes did men turn to the statistical method? This question has been very ably discussed by Theodore Merz in the second volume of his "History of European Thought in the Nineteenth Century," and we cannot do better than follow his development of the matter. Speaking of the origin of statistics, Merz says (*loc. cit.*, pp. 554–555):

"That which everywhere oppresses the practical man is the greater number of things and events which pass ceaselessly before him, and the flow of which he cannot arrest. What he requires is the grasp of large numbers. The successful scientific explorer has always been the man who could single out some special thing for minute and detailed investigation, who could retire with one definite object, with one fixed problem into his study or laboratory and there fathom and unravel its intricacies, rising by induction or divination to some rapid generalization which allowed him to establish what is termed a law of general

aspect from which he could view the whole or a large part of nature. The scientific genius can 'stay the moment fleeting'; he can say to the object of his choice, 'Ah, linger still, thou art so fair'; he can fix and keep the star in the focus of his telescope, or protect the delicate fiber and nerve of a decaying organism from succumbing to the rapid disintegration of organic change. The practical man cannot do this; he is always and everywhere met by the crowd of facts, by the relentlessly hurrying stream of events. What he requires is grasp of numbers, leaving to the professional man the knowledge of detail. Thus has arisen the science of large numbers or statistics, and the many methods of which it is possessed."

Further on the same author says of the origin of the science of probability (*loc. cit.*, pp. 567–568):

"The necessity of having recourse to elaborate countings, to registrations of births, deaths, and marriages, to lists of exports and imports, to records of consumption and production of foodstuffs and many other items, forced upon those who were intrusted with the gathering and using of these data the observation that all such knowledge is incomplete and inaccurate. Owing to the variability, within certain limits, of recurring events and the errors of counting and registration, we have to content ourselves always with approximation instead of certainty. Error bulks

very largely in all statistics, and vitiates them; and as regards coming events, our minds are in a state of expectation rather than of assurance. But events can be more or less probable, errors can be greater or smaller, cumulative or compensatory, and our expectations may be well- or ill-founded. And so there has arisen the science of Probabilities and of Chances, and the Theory of Error, two subjects intimately interwoven. The former arose in the seventeenth century out of the frivolous or vicious practice of betting and gambling, whilst the latter was founded when astronomical observations accumulated, and the question presented itself how to combine them so as to arrive at the most reliable result."

Now from these two quotations, which may certainly be considered as fairly stating the case, it is apparent that those circumstances which led men to turn to statistical methods of reasoning and investigation were not such as grow out of an increasing precision and certainty of knowledge about the events or things under consideration, but rather were quite the opposite. In other words, the statistical point of view, in the first instance, was adopted as an admittedly imperfect means of getting some sort of knowledge about a class of events concerning which it was difficult or impossible to get by other methods the precise or particular kind of knowledge which was wanted.

## III

A careful consideration of the history of statistical science, as well as of the present day application of these methods, leads to the conclusion that statistical methods are used for two sorts of purposes, or to gain two sorts of knowledge about events or things.

*A*. On the one hand the statistical method finds one of its chief uses in furnishing a method (and the only one known in science) of describing a group in terms of the group's attributes, rather than in terms of the attributes of the individuals which compose the group. This important use of the statistical method is not usually formulated in just the way I have put it here, or in an earlier chapter (p. 54). On that account it may be well to consider the point a little more analytically.

What sorts of positive, definite, and exact knowledge do statistics give us?

1. Precise knowledge of the *composition* of groups or masses. This is the knowledge gained by *counting*. Suppose we find a basket containing a number of balls of several different colors, and proceed to count them with the following results:

7 Reds
9 White
2 Black
1 Green

Such a count furnishes us at once with a great deal of perfectly definite and precise information about this group or population of balls. For example, the count tells us that it will never be possible to draw more than one pair of balls of which one member is green. This is a definite attribute of this population which may be used to differentiate it from other populations. In this particular population only one green ball occurs.

This sort of knowledge derived by counting is perfectly definite and precise so far as relates to the particular group or mass which it concerns in any particular case. It does not involve any approximation, or probability, and is as precise as knowledge of the individual. It, however, pertains to the group. It forms a part of a proper scientific description of a group.

2. Knowledge of certain *abstract qualities* of groups or masses. This knowledge is obtained by calculation from the counted data. The more important of the abstract qualities of groups are:

*a.* The *center* or *typical condition* of the group; or the condition about which the individuals composing the group cluster. This is variously measured: by the arithmetic mean, which gives the center of gravity of the group, by the median, which tells the point on either side of which exactly half the individuals fall, by the mode, which tells the point of greatest frequency of occurrence in the group, etc.

*b.* The degree of *individual diversity* comprised in the group. This attribute, called the variability of the group, is again variously measured: by standard deviations, coefficients of variation, etc.

*c.* The *degree of symmetry* of the distribution of the individuals composing the group. This is measured by the skewness or other related constants.

*d.* Various other attributes of distributions might be here included, such as, for example, the kurtosis, but for purposes of the present general analysis this scarcely seems necessary. Though some of these attributes involve very complex mathematical expressions for their measurement, the general fact remains clear that they are all attributes of groups or masses which are described by the statistical constants.

One point here we must be quite clear about. This is that the kind of knowledge discussed under this heading 2 is just as definite and precise, and involves as little approximation and indeterminism, as does any piece of individualistic knowledge, *so long as we confine our attention solely to the particular group discussed in a particular single case.* We are accustomed to stating means, for example, with probable errors. But this is only because it is proposed to extend the conclusions beyond or outside of the particular group and the particular instance for which the mean was calculated. *For*

*that group and that instance* the mean is perfectly exact and precise to that degree of precision denoted by the unit of measure used, assuming that no arithmetical mistakes have been made in its computation. Thus suppose one measures the stature of three men to the nearest inch, and then calculates the average. The result is, without any probable error, the average height, at the particular moment when they were measured, of *those three men exact to the unit of measurement used.* It describes and measures precisely an attribute of those men considered as a group. But if we were to consider this result from the viewpoint of whether it gave a reasonable measure of the average height of men in general, or from the viewpoint of whether it gave a proper value for the mean height of these men when repeatedly measured under varying conditions, it would clearly be subject to a large probable error. It would, in point of fact, have lost its character of precise and definite knowledge, and have become a more or less poor approximation.

3. Precise knowledge of the *degree of association* or contingency between different events or characters within a group. This is furnished by the method of correlation in one or another of its various forms. By this general method we are able to measure precisely the degree of resemblance between the individuals composing a group in respect to one or more characters. So long as attention

is confined to the particular group on which the measurement is made, and to that group alone, and to a single instance (in time) the knowledge gained is precise. It is a part of the description of the attributes of that group. When we pass from that particular group to other groups or individuals our results are no longer precise, but inferential, and the probable errors tell us something about the degree to which the inference is trustworthy.

Summarizing the results of the above analysis, we see that the statistical method can

1. Furnish precise descriptive knowledge about groups. This knowledge is of various sorts. It is definite and precise so long as attention is confined solely to the particular group and the particular instance on which it is based.

2. The knowledge gained by the statistical method, as we have analyzed it above, precise though it may be, *pertains to the group and not to the individual.* It is exact knowledge about the composition, or attributes, or contingencies of masses or groups.

3. This ability to describe groups in terms of the groups' own attributes, which is an unique property of the statistical method, is extremely useful in the practical conduct of scientific investigations. It makes the statistical method an absolutely essential adjunct to every other scientific method, and particularly to the experimental.

This fact is just now beginning to be recognized by some experimentalists and hailed as a rather original thought. It is not new.

*B.* We may turn now to a wholly different aspect of the statistical method, wherein it is used for the purpose of predicting or estimating the probable or the approximate condition in the *individual* from a statistical examination of the condition in the mass or the group. Resort is had to the statistical method for this purpose primarily in those cases where the outcome of the event, or the condition of the thing, is determined by the combined action of a large number of small causes, each about equally influential upon the final result.

Originally the statistical method was only employed for this second purpose in cases where, because of the multiplicity of the cause groups involved in the determination of the event, and the consequently small effect of each, it was impossible to make any reasonable prediction regarding an individual from an examination of that individual alone. Such employment might be considered legitimate, though not very fruitful, on the ground that prediction so made, uncertain and doubtful as it may be, is after all perhaps better than no prediction at all. As time has gone on, however, there has been an increasing tendency to assume that this use of the statistical method had general *a priori* validity and could be profitably

employed in all sorts of cases. This point of view reaches, it seems to me, its limit in the sentence already quoted from Royce, which I venture to repeat with one word italicized. "There is, therefore, good reason to say that not the mechanical but the statistical form is the canonical form of scientific theory, and that if we knew the natural world millions of times more widely and minutely than we do, the mortality tables and the computations based upon a knowledge of averages, would express our scientific knowledge about *individual* events, much better than the nautical almanac would do."

This leads us to consider carefully the general question of the validity on the one hand, and the usefulness on the other hand, of this whole second mode of employment of the statistical method. It is the one which has attracted the greatest attention because of its essentially spectacular nature coupled with a sort of mysteriousness bordering upon the miraculous. It seems a wonderful, indeed almost a superhuman, accomplishment to be able to say in the manner of the oracles of old, "So many men will commit suicide next year."

## IV

Since Clerk-Maxwell introduced statistical modes of reasoning into physical science there has been an ever increasing tendency to regard the

universe as organized on a statistical plan. This has come to carry with it two implications, one of which is quite fallacious and the other partly so.

The first of these is that the individual events, of which all the causes are not precisely known to us, are indeterminate. Such an assumption is of course unwarranted. Because we do not know all the causes leading to a particular event does not mean that that event is any the less precisely determined by the course of antecedent events. Consider a box containing 100 consecutively numbered cards. Suppose one card were to be drawn and that it bore the number 36. It would be quite impossible to formulate precisely all the causes which led to the drawing of the number 36 on the particular occasion considered, but it is equally impossible to conceive that this result was not definitely "caused." In other words, there clearly was a whole train of antecedent circumstances, which taken all together definitely resulted, *and could only have resulted*, in the drawing of the number 36. The too prevalent conclusion that the application of the statistical method or statistical modes of thought implies phenomenal indeterminism in the individual case is totally fallacious.

The second currently accepted implication of a statistical view of the universe is that in general a particular event or phenomenon is the outcome

of the combined action of a great number of causes, each of which alone produced but a small part of the final total effect. There is clearly so much truth in this point of view as is included in the fact that individual events or phenomena do, in some degree or other, vary, and further these variations in general distribute themselves more or less in accord with the well-known laws of errors. But the assertion that events are individually the outcome of the action of great numbers of causes, each of which had a small part and a part significantly equal to that played by every other one of the causes concerned in the final result, is only true if the "universe of discourse" is indefinitely extended in time. But *practically* science works in a definitely and rather narrowly limited universe of discourse so far as concerns time. One of the causes for the writing of these lines is that a certain worthy was not shipwrecked in voyaging to this country nearly 300 years ago, since if he had been shipwrecked presumably I should not exist and therefore could not write these words. But practically this cause had very little to do with determining that I, being here in existence, should write this book rather than do various other things which I might have done instead. It undoubtedly is true that a vast number of small causes do play a part in the determination of any particular event. But, in many of the events, at least, in which science is

interested, these multitudinous minor causes do not play any significant part in the differential determination of a particular event at a particular instant of time. There is in connection with the causation of most events some one or two, or at most a very few, outstanding cause groups which, for all practical purposes, at a given moment completely determine their occurrence. The total effect of all the vast number of other minor causes concerned in the remote past is so minute, as compared with the part played by the really determinative ones at the moment, as to be negligible. In other words, all natural cause groups are not small, nor of equal (balanced) values in the final determination of the event to which they relate, provided we confine ourselves to the time limits of finite practical operations. Yet something very much of that sort seems to me to be implicitly involved in any such generalized statistical view of the universe as that which Professor Royce holds up for our admiration as a sort of scientific ideal.

The fact that all natural causes or cause groups are not equally significant quantitatively is, of course, what makes the experimental method fruitful — one might even say possible — in science. The very essence of the experimental method is that the conditions for the happening of an event are so arranged that the influence of one putative causal factor may be tested at a

time. If with a radical change in this one factor, whilst all others remain, so far as may be, constant, no change in the happening of the event is observed, the experiment has shown that this particular factor has no *significant* causal relation to the happening of the event. If a marked change in the happening of the event is observed always to follow the change of conditions of operation of the factor under investigation, then clearly this factor plays a determinative part.[1] In other words, it is a fundamental logical prerequisite of the experimental method if it is to be successful (that is, contribute to knowledge) that it operate in a universe in which all causal factors are not of equal quantitative significance at any given instant of time.

Clearly experimental analysis of this sort would have quickly discovered, if the common sense of men had not long previously shown, that the course which a particular event is going to take is not immediately the result of the action of an indefinitely large number of individually insignificant causal factors, but that it is the outcome of the action of a few immediately *determinative* factors and the effect of the indefinitely large number of historically antecedent small causes is insignificant in the sense of being differential. Generalized, the point may be put in

[1] Cf. Jennings' valuable paper on radical experimental analysis already referred to in an earlier chapter (p. 13, *supra*).

this way: an event *A* is about to happen. It may happen in any one of *n* different ways, each one of which ways may be designated by a letter, *l*, *p*, *r*, *t*, etc. Now an indefinitely large number of causes are concerned in bringing it about that the event *A* is going to happen, and that it can equally well happen either as *l*, *p*, *r*, *t*, etc. In other words, the setting of the stage for the event has involved a vast number of small and balanced causes. But the causes which are differential in the particular case, that is, which determine that *A* shall happen in the *p* way this particular time, and not in the *l*, the *t*, or any other way, are, in general:

1. Few in number.
2. Immediate in time.
3. Large in relative quantitative effect.

The point under discussion may perhaps be made plainer by a homely illustration. Suppose a man steps up behind a mule and prods the creature with his walking stick. The human intellect is unequal to the task of predicting exactly, in the particular case, what precise portion of the man's body the mule's hoof will land upon. A multitude of minor causes will affect this: the relative height of the man and the mule, the age of each, the place poked with the walking stick, the degree of fatigue of the mule, the temperature, the season of the year, and countless other things

have an influence in determining just the precise spot where the mule's foot and the man's body come together. These could be investigated statistically and tables drawn up from which one could predict the part of the man which would most probably receive the hoof. But what a silly, futile piece of business this all would be, since clearly the influence of all of these small causes on what happens to the man is stupendously overshadowed by the results of two factors; namely, putting himself behind a mule and prodding the animal with a stick. Of course, a vast number of antecedent causes are involved in the setting of the stage, but these are not differential in the determination of the end event of the series.

The preceding illustration has nothing directly to do with science, but the essential point involved operates in the use of the statistical method as a weapon of scientific research. This method being, as we have seen elsewhere, only a descriptive method, it cannot, any more than any other descriptive method, tell us anything *directly* about the causes involved in the determination of any events or phenomena under consideration. It may be of great aid, in combination with the experimental method, in helping to arrive at such knowledge, but alone and of itself it cannot directly furnish knowledge of causes of individual events. Yet the statistical method, particularly in that phase of it which we have here under dis-

cussion, which essays to predict the probable condition of the individual from the knowledge of the mass, *seems* to furnish information about causes. It wears a specious air of bringing a kind of knowledge which in reality it not only never does, but from the very nature of the case never can furnish.

## V

Let us consider now a little more in detail the nature of the prediction of the probable condition of the *individual* from a knowledge of the *mass* or group. It has been shown in an earlier section (p. 83) that statistics give perfectly definite and precise, and often very useful knowledge about masses or groups. We are now, however, not concerned with this as group knowledge, but rather with one use to which such knowledge has been put. This use is that which is comprised in the subject of statistical probabilities, and which involves the drawing of conclusions as to the *probable* condition of the individual, based on an *exact* knowledge of the mass.

'In order to approach the subject in the simplest way let us consider a concrete case. Suppose a problem of the following sort were to be set before us for answer: What is the probability that, at some chosen moment of time, the next birth to occur in, let us say, the city of Baltimore, will be of a white child. Now if we look at this as a

question in statistical probability the appropriate way, of course, to go about solving it is to turn up the registration reports for the city of Baltimore covering a period of years, and find out what is the proportion of white to colored births in that city. Then, by the simplest theorem in the calculus of chance, the probability that the next birth will be of a white child will be given by a fraction of which the numerator is the number of white children born in Baltimore and the denominator is the total number of children born in Baltimore, both figures including the same period of time. The difference between the fraction so obtained and 1 will be the probability that the next birth will be of a child not white; that is, colored. When we have obtained such a fraction we have a definite piece of statistical knowledge, but of just what use is it so far as concerns the individual case? It implies no biological knowledge of any kind; no knowledge of the laws of heredity. It really adds essentially, it seems to me, to the sum total of the world's knowledge only one thing. That thing is the proper betting odds on what the color of the next child born in the city will be. This knowledge would really be useful, in a pragmatic sense, only provided some one wishes to gamble upon that event.

Of course the statistical count, on which the probability is based, in itself furnishes definite and precise information about the population of

Baltimore, *as a population*. This may be useful. What we are now considering, though, is knowledge about individual cases.

Let us see what a totally different kind of ability to predict the future event in an individual case is gained when we take into account one single biological fact of an individualistic instead of a statistical character. Suppose, that is to say, that we are informed that the mother of the next baby to be born in Baltimore is black. It needs no argument to show how much more precise is our prediction as to the color of the next baby under these conditions.

This illustration brings out clearly the difference between the two possible bases for the prediction of a future event. On the one hand, such prediction may be based on statistical ratios. This means merely a count of an indefinitely large past experience regarding the occurrence or failure of the event, but in no way takes into account the causes which underlie the happening of the event in any particular case. On the other hand, we have the prediction which is based on a definite knowledge of the determinative causes which bring about the happening of a particular individual event of the sort in which we are interested and about which we are to predict. There can be, it would seem, no comparison between the usefulness, in the pragmatic sense, of these two kinds of knowledge. The statistical knowledge on which

a statistical prediction is made is essentially the most sterile kind of knowledge that one can possibly have so far as concerns the individual event. It merely gives one the betting odds for or against the occurrence of an event, and absolutely nothing more. Now a wager, however large, in the scientific sense neither discovers, expounds, nor is a criterion of the truth. Bets, in other word, are not evidence, though the statistician sometimes seems to forget this, and to deal with statistical ratios as though they had probative worth in regard to phenomena.

On the other hand, a prediction based on experimentally acquired knowledge of the determinative cause of the individual event brings with it a real knowledge of a natural phenomenon. The predictions so made may not always turn out correct, but when they do not, it incites us to investigate the particular disturbing factor which under such circumstances may overwhelm the normally determinative cause of a particular event.

## VI

*Man soll das Kind nicht mit dem Bade verschütten.* The critical reader may be inclined to think that this is exactly what the discussion in the preceding sections has done. If, as has there been suggested, that part of the statistical method which uses the calculus of probability as a basis for the

prediction of future events, gives only a knowledge of betting odds, one may ask: what about the whole concept of probable error? The value of this concept in scientific research is unquestioned (cf. p. 59, *supra*). Yet plainly the whole concept has its basis in the calculus of probability. Has not our discussion led us unwittingly into a serious contradiction?

I think not. Let us examine the probable error concept a little more carefully than we have yet done. Suppose we read that the mean length of the thorax of a thousand fiddler crabs is 30.14 ± .02 mm. Just what does this actually mean? Accepting the figures at their face value, or, put another way, assuming that the mathematical theory on which the probable error was calculated was the correct one, the figures mean something like this: If one were to take, quite at random, successive samples of 1000 each from the total population of fiddler crabs and determine the mean thoracic length from each sample, these means would all be different from each other by varying amounts. In other words, no single sample would give us the absolutely true value of the mean thoracic length of the whole fiddler crab *population.* This true value is in an absolute sense unknowable, because, for one reason, always we must come at the finding of it by the way of random sampling, and sampling means variation. Now it is an observed fact of experience that the

variations due to random sampling distribute themselves according to a definite law of mathematical probability. Knowing this law, it is clearly possible to state the mathematical probability for (or against) any particular deviation or variation occurring as the result of random sampling. Exactly this is what the probable error does. It says, in the particular case here considered, that it is an even chance, that a deviation or variation in the value of the mean as great as or greater than .02 mm. above or below will occur as a result of random sampling. Or, put in another way, if we took successive samples of 1000 each from this crab population, it is an even bet that the value of the mean from any sample would fall between $30.14 + .02 = 30.16$, and $30.14 - .02 = 30.12$.

Now all the knowledge that this probable error furnishes is this: that if a man were to say, "I'll bet a thousand dollars that the mean thoracic length of the next sample of fiddler crabs you measure will be either over 30.16 mm. or under 30.12 mm.," one would not be justified in offering odds. He could wager on even terms. Either party involved in the transaction would be as likely to lose (or to win) as the other.

Putting the case in this way, it is clear that this is the same kind of knowledge which comes from an examination of probable errors as that discussed in the preceding section. It is a knowledge

of betting odds. It has no necessary relation *per se* to any physical, chemical, or biological laws. It merely informs one how he may safely gamble on an event if he is so minded and can find some one else ready to do the same thing.

Wherein lies the value of the probable error concept for science, then? Simply in that it serves as a test or check on every mode of research in science. So far as I can see, the calculus of probability, in and of itself alone, is not and never can be an effective weapon of research for the discovery of truth in phenomenal science, be it physical or biological. Yet it operates as an ever-present test of the trustworthiness of the results obtained by modes of research which are in themselves adapted to making discoveries about phenomena. The student of probability says something like this to the experimentalist: "Yours is the way to find out the significant underlying causes of phenomena. Let it be practiced with all zeal, but let it be remembered that you operate in a finite way in a finite universe, and consequently all your results are subject to such fluctuations and variations as experience has shown arise from random sampling. I regret that I cannot directly and alone discover significant causes, but at any rate I can furnish you a test whereby you may reasonably judge whether your result is significantly influenced by these fluctuations of random sampling."

## VII

To sum the whole matter up: I have tried to show that the statistical method in science has been used to do two things.

The first of these is a unique function of the method — to furnish a description of a group of objects or events in terms of the group's attributes rather than those of the individuals composing the group. Herein lies the great value of the statistical method. It is, however, a descriptive method only and has the limitations as a weapon of research which that fact implies.

The second purpose that the statistical method has been called upon to accomplish is the prediction of the individual case from a precise knowledge of the group or mass. This involves something really additional to the statistical method *per se;* namely, the mathematical theory of probability. We have seen that this side of the statistical method gives only a somewhat sterile kind of knowledge so far as concerns individuals; namely, a knowledge of betting odds. The theory of probability grew up about the gaming table, not in the laboratory. Its place in the methodology of science is not an independent one. By it alone one cannot discover new truths about phenomena. But it is a highly important adjunct to other modes of research.

Plainly, however, one cannot regard statistical

knowledge in general as a higher kind of knowledge than that derived in other ways. Nor is the statistical method to become the dominant or exclusive method of science, though it will always be useful, and in many fields an essential method. It will find its chief usefulness, first in its sphere of furnishing shorthand descriptions of groups, and second in furnishing a test of the probable reliability of conclusions.

# CHAPTER IV

## CERTAIN LOGICAL AND MATHEMATICAL ASPECTS OF THE PROBLEM OF INBREEDING [1]

### I. The Analysis and Measurement of Inbreeding

The effect of inbreeding on the progeny is a much-discussed problem of theoretical biology and of practical breeding. It has been alternately maintained, on the one hand, that inbreeding is the most pernicious and destructive procedure which could be followed by the breeder, and on the other hand, that without its powerful aid most of what the breeder has accomplished in the past could not have been gained and that it offers the chief hope for further advancement in the future. While there is now, among animal breeders at least, a more widespread tendency than was formerly the case towards the opinion that inbreeding *per se* is not a surely harmful thing, nevertheless this opinion is by no means universally held, and in any case does not rest upon a definite and well-organized body of evidence.

[1] This paper is based upon a series of "Studies on Inbreeding" which have recently been published, chiefly in the *American Naturalist*, during 1913, 1914, and 1915.

Aside from a relatively small amount of definite experimental data one's judgment is finally formed on the basis of his interpretation of the vast accumulation of material comprised in the recorded and unrecorded experience of the breeders of registered (pedigreed) livestock.

This material recorded in the books of registration far exceeds in amount and in diversity any which could possibly be obtained experimentally on the same forms of life. It must be said, however, that the discussion of it with a view to an analysis of the effects of inbreeding, though undertaken at greater or less length by a number of men including Lehndorff, von Oettingen, Bruce Low, Hoesch, Chapeaurouge, Bunsow, Strang, and others, has not led to results characterized by the precision, the definiteness, or the quality of getting at fundamentals demanded in the present state of the science of genetics.

The lack of precision and fundamental character in the studies alluded to is not primarily to be attributed to any inherent defect in the material. In the breeding of all of the domestic animals inbreeding has been practiced; in many instances to a very marked degree. Further, the manner in which the inbreeding has been done (the types of relationship-matings) exhibits a most intricate diversity, from which different types may be picked out for analysis in any reasonable quantity. The records are accurate, within their limitations,

to a high degree. Probably no experimentalist's records of descent are more accurate, considering the relative numbers involved in the two cases.

The real need, I venture to think, has been for an appropriate and valid method of pedigree analysis, which possessed generality, and could on that account be depended on to give comparable results when applied to two (or more) different pedigrees. Specifically, there seems not to have been worked out any *adequate general method of measuring quantitatively the degree of inbreeding which is exhibited in a particular pedigree.* Without such a measure it is clearly impossible to proceed far in the analysis of the kinship aspect of inbreeding.

It is the purpose of this paper to present a method for measuring and expressing numerically, in the form of coefficients, the degree of inbreeding which exists in any particular case. I shall endeavor to show that the method is (*a*) *unique*, in the sense that the values obtained in any particular instance can only be affected by the degree or amount of inbreeding which has been practiced in the line of descent under consideration, and (*b*) *general*, in the sense that it is equally applicable to all pedigrees and to all degrees and types of inbreeding.

## Preliminary Definitions

In attempting any general analysis of the problem of inbreeding from the theoretical stand-

point one is confronted with the necessity for a definition of inbreeding which shall be at once precise and general, that is, such as to include all of the many diverse ways in which this sort of breeding may be practiced. A great number of definitions of the concept "inbreeding" have been proposed in the literature of genetics. I shall not attempt to review these definitions here, since to do so would serve no useful purpose in the present connection. A careful consideration of them is bound, I think, to lead one to the conclusion that they have been, in general, based on grounds of practical expediency rather than critical biological analysis.

Clearness and simplicity of thinking will be gained by approaching the problem *de novo*. Leaving aside for the moment all consideration of details as to how a particular piece of inbreeding is done, it is clear that underlying all definitions of inbreeding is to be found the concept of a *narrowing* of the network of descent as a result of mating together at some point in the network individuals genetically *related* to one another in some degree. Let us take this as our basic concept of inbreeding. It means that the number of potentially *different* germ-to-germ lines or "blood-lines" concentrated in a given individual animal is fewer if the individual is inbred than if it is not. In other words, *the inbred individual possesses fewer different ancestors in some particular generation or gen-*

*erations than the maximum possible number for that generation or generations.* This appears to be the most general form in which the concept of inbreeding may be expressed.[1] In whatever way the mating of relatives is accomplished, or whatever the degree of relationship of the individuals mated together, the case in last analysis comes back to the above statement; namely, that there are actually in the pedigree of the inbred individual fewer *different* ancestors in some particular generation or generations than the maximum possible number.[2]

The idea suggested in the foregoing paragraph may be expressed symbolically as follows. If there is absolutely no collateral relationship between any of the individuals in a pedigree, the number of different individuals in succeeding ancestral generations will be given by the series

$$x \leftrightarrow (1)\ 2 \leftrightarrow (2)\ 4 \leftrightarrow (3)\ 8 \leftrightarrow (4)\ 16 \leftrightarrow (5)\ 32 \leftrightarrow (n)\ 2^n, \qquad \text{(i)}$$

where the numbers in parenthesis denote the numbers of the ancestral generations (1 = parents,

[1] This, of course, looks at the matter primarily from the standpoint of kinship. This is all that is intended here.

[2] This generalized concept of inbreeding seems to me to be in essential (though not entirely in verbal) agreement with that of O. F. Cook, whose interesting general discussions of this and related problems are summarized in a recent paper ("The Superiority of Line Breeding over Narrow Breeding," U.S. Dept. Agr., Bur. Plant Ind. Bul. 146, 1909). I use "inbreeding" as a generic term, while Cook regards it as a species of "line breeding." This seems to me to be a purely terminological difference, and not of great consequence.

2 = grandparents, 3 = great-grandparents, and so on), and the free figures denote the maximum possible number of different ancestors to the indicated generation. If in any generation in the series relatives are bred together, the same individual will appear more than once in the ancestral series, and the number of different individual ancestors in the higher terms will be accordingly diminished below the maximum number as given in (i). The series will then become

$$x \leftrightarrow (1)2 \leftrightarrow (2)4 - y_1 \leftrightarrow (3)8 - y_2 \leftrightarrow (4)16 - y_3 \leftrightarrow (5)32 - y_4 \ldots, \qquad \text{(ii)}$$

where $y_1, y_2, y_3, \ldots$ may, in the $n$th generation, have any value not greater than $2^n - 2$, in the case of organisms in which two individuals must coöperate in the process of reproduction. The final limiting case is, of course, self-fertilization, where the number of ancestors reduces to 1 in each generation.

## The Measurement of the Degree of Inbreeding

This brings us to a consideration of a practical and general measure of the degree of inbreeding exhibited in a particular pedigree. This problem has been attacked by a number of investigators, but so far as I have been able to learn all previous measures have been modifications in one form or

another of the scheme of Lehndorff. This plan[1] took account, as a measure of inbreeding, only of the number of generations intervening between that generation in which relatives were bred together, and that generation in which their first common ancestor was found. Thus Lehndorff says:[2]

"I am of opinion, that a horse should only be termed *in-bred*, when in sum total less than four degrees lay between its parents and their common ancestor; in other words, when the children or grandchildren of a stallion or a mare are mated, I call their produce *in-bred;* but this term does not apply to the produce of great-grandchildren of the common ancestor. We must not forget that in the pedigrees of horses the word brother or sister often means half-brother or half-sister, and that here the definition borrowed from the human family connection is not applicable.

"As breeding within *moderate relationship* I reckon the mating of stallion and mare that are removed from their common ancestor four, five, or six degrees. It is indifferent whether they are on both sides equidistant from, or one of them nearer to the male or female progenitor than the other."

Von Oettingen used a measure exactly the same in principle as this of Lehndorff's. The system of

[1] Cf. Lehndorff, G. "Horse-breeding Recollections," Philadelphia, 1887.

[2] *Loc. cit.*, p. 49.

Bruce Low, though somewhat differently stated, comes to essentially the same thing, so far as I am able to determine from abstracts, this author's original writings not having been accessible to me.

All systems based on the number of "free generations" alone do not furnish a precise or reliable measure of the real intensity of inbreeding. The essential reason for this failure, stated baldly, is that they do not take account of the composition of the generation to which the "common ancestors" of an inbred pair belong.

In developing a general measure of the intensity of inbreeding we may well start from the conception set forth in the preceding section; namely, that the inbred individual possesses fewer different ancestors than the maximum possible number. Besides this factor account must be taken of the generation or generations in which the reduced number of different ancestors is found, and the extent to which these generations are removed (in the sense of Lehndorff discussed above) from the individual or generation under consideration. In other words, the two factors which must be included in a general measure of the intensity of inbreeding are (*a*) the *amount* of ancestral reduction in successively earlier generations, and (*b*) the *rate* of this reduction over any specified number of generations.

Both of these demands are met, I think, by taking as a measure of the intensity of inbreeding

in any generation the proportionate degree to which the actually existent number of different ancestral individuals fails to reach the maximum possible number, and by specifying the location in the series of the generation under discussion.

The statement is amplified and made more precise in the following propositions:

1. The production of the individual must be the point of departure in any analytical consideration of inbreeding, leading towards its measurement. That is, the question to which one wants an answer is: What degree of inbreeding was involved in the production of this particular animal?

2. It is therefore necessary practically to *start* with the individual and work *backwards* into the ancestry in measuring the inbreeding, rather than to start back in the ancestry and work down towards the individual.

3. In the genetic passage from the $n + 1$th generation to the $n$th, or in other words the contribution of the matings of the $n + 1$th generation to the total amount of inbreeding involved in the production of an individual, the degree of inbreeding involved will be measured by the expression,

$$Z_n = \frac{100\ (p_{n+1} - q_{n+1})}{p_{n+1}}, \qquad \text{(iii)}$$

where $p_{n+1}$ denotes the maximum *possible* number of different individuals involved in the

matings of the $n + 1$ generation, and $q_{n+1}$ the *actual* number of different individuals involved in these matings. $Z_n$ may be called a *coefficient of inbreeding*. If the value of $Z$ for successive generations in the ancestral series be plotted to the generation numbers as a base, the points so obtained will form a curve which may be designated as the *curve of inbreeding*.

It will be noted that the coefficient of inbreeding $Z$ is the percentage of the difference between the maximum possible number of ancestors in a given generation, and the actual number realized, in the former. The coefficient may have any value between 0 and 100. When there is no breeding of relatives whatever (that is, in the entire absence of inbreeding), its value for each generation is 0. As the intensity of the inbreeding increases, the value of the coefficient rises.

4. The above measure of inbreeding has to do primarily with the *relationship* aspect of the problem.

5. Since the only possible infallible criterion of relationship between individuals is common ancestry in some *earlier* generation, we are led to the practical rule, in measuring the degree of inbreeding in a pedigree, to regard all different individuals as entirely unrelated until the contrary is proved by the finding of a common ancestor. This no doubt appears at this stage of the discussion as an exceedingly obvious truism. The reader is urged

to accept it as such, and hold fast to it, because it will help him over some apparent paradoxes later.

The method of calculating coefficients of inbreeding, and their real significance, will be made much clearer by the consideration of illustrative examples of their application. To these we may therefore turn.

## The Calculation of Coefficients of Inbreeding

We may first consider some simple hypothetical pedigrees, before attacking the more complicated ones actually realized in stock-breeding.

### Illustration I. Continued Brother × Sister Breeding

Let us begin with the most extreme type of inbreeding possible; namely, the mating of brother with sister for a series of generations. Pedigree Table III gives the pedigree of an individual so bred.

Let us now proceed to the calculation of the coefficients of inbreeding, $Z_0$, $Z_1$, $Z_2$, and $Z_3$. For $Z_0$ we have

$$p = 2,$$
$$q = 2,$$

PEDIGREE TABLE I. (HYPOTHETICAL)

TO ILLUSTRATE THE BREEDING OF BROTHER × SISTER, OUT OF BROTHER × SISTER, CONTINUED FOR A SERIES OF GENERATIONS

| | | | | |
|---|---|---|---|---|
| | | | $e$ / $f$ | $g$ / $h$ |
| | | $c$ / $d$ | | $g$ / $h$ |
| | $a$ | | $e$ / $f$ | $g$ / $h$ |
| $x$ | | | | $g$ / $h$ |
| | | $c$ / $d$ | $e$ / $f$ | $g$ / $h$ |
| | $b$ | | | $g$ / $h$ |
| | | | $e$ / $f$ | $g$ / $h$ |
| | | | | $g$ / $h$ |
| Ancestral Generation . . | 1 | 2 | 3 | 4 |

whence

$$Z_0 = \frac{100\ (0)}{2} = 0.$$

In the same way,

$$Z_1 = \frac{100\ (4-2)}{4} = 50.$$

$$Z_2 = \frac{100\ (8-2)}{8} = 75.$$

$$Z_3 = \frac{100\ (16-2)}{16} = 87.5.$$

These results may be expressed verbally in the following way: In the last two ancestral generations $x$ is 50 per cent inbred; in the last three generations it is 75 per cent inbred; and in the last four generations it is 87.5 per cent inbred.

This pedigree table and the constants will repay further consideration, since the case is a limiting one. With the table at hand it is possible to grasp a little more clearly the precise meaning of the coefficients of inbreeding. Thus it is seen that what the value of $Z_1 = 50$ really signifies is that because the individuals $a$ and $b$ were brother and sister the number of different ancestors which $x$ can possibly have in any ancestral generation cannot be *more* than 50 per cent of the total number theoretically possible for the generation. That is, $x$'s sire and dam having been brother and sister means that $x$ cannot have had more than 2048 different great-great-great-great-great-great-

great-great-great-grandparents, instead of the possible 4096. He may have had fewer than 2048, but $Z_1 = 50$ tells us that he could not have had more. Similarly, $Z_2 = 75$ indicates that since $c$ and $d$, the grandsire and granddam of $x$, were brother and sister, $x$ cannot have in any earlier ancestral generation more than 25 per cent of the theoretically possible number of ancestors for that generation. And so on for the other values of $Z$.

In the limiting case of the closest inbreeding possible the successive $Z$'s will have the values given in the table on the opposite page.

From this table it is apparent that while the narrowing or exclusion of the possible different source lines of descent proceeds very rapidly in the first few generations of brother × sister breeding, only relatively little change is made by further generations of this sort of breeding. Thus in seven generations of brother × sister breeding all but about 1.5 per cent of the potentially different ancestral "blood lines" will have been eliminated. After 16 generations of this sort of breeding (a number easily attainable in ordinary breeding experiments) an individual so bred can by no chance possess more than $\frac{3}{1000}$ of one per cent of the different lines of ancestral descent which are theoretically possible. This table strongly suggests that if, in an experiment to test the influence of inbreeding, no particular effect is observed

TABLE 1

VALUES OF THE SUCCESSIVE COEFFICIENTS OF INBREEDING ($Z_0$ to $Z_{15}$) IN THE CASE OF THE MOST INTENSE INBREEDING POSSIBLE (BROTHER × SISTER OUT OF BROTHER × SISTER — CONTINUED)

| COEFFICIENT OF INBREEDING | ANCESTRAL GENERATIONS INCLUDED | NUMERICAL VALUE OF COEFFICIENT |
|---|---|---|
| $Z_0$ | 1 | 0 |
| $Z_1$ | 2 | 50 |
| $Z_2$ | 3 | 75 |
| $Z_3$ | 4 | 87.5 |
| $Z_4$ | 5 | 93.75 |
| $Z_5$ | 6 | 96.875 [1] |
| $Z_6$ | 7 | 98.4375 |
| $Z_7$ | 8 | 99.21875 |
| $Z_8$ | 9 | 99.609375 |
| $Z_9$ | 10 | 99.8046875 |
| $Z_{10}$ | 11 | 99.90234375 |
| $Z_{11}$ | 12 | 99.951171875 |
| $Z_{12}$ | 13 | 99.9755859375 |
| $Z_{13}$ | 14 | 99.98779296875 |
| $Z_{14}$ | 15 | 99.993896484375 |
| $Z_{15}$ | 16 | 99.9969482421875 |

during ten generations of brother × sister breeding, it is extremely improbable that any effect will be

[1] There is, of course, no further point in the retention of all the decimals in these coefficients than to make plain the law of their formation. In the case of maximum inbreeding here illustrated, the $Z$'s are the successive terms of a series in which any term is equal to the preceding term plus one half the difference between the preceding term and 100, or in which the successive differences are halved.

produced by a further continuation of the same method of breeding. If an apparent effect should suddenly appear some time later than the tenth generation, the case would need the most critical scrutiny, to determine whether the observed effect had really been due to the inbreeding, rather than to some other unsuspected cause.

The values of the $Z$'s in Table I are maxima. No particular coefficient of inbreeding can have a higher value than that given in the table. It is not possible, for example, so to breed any animal

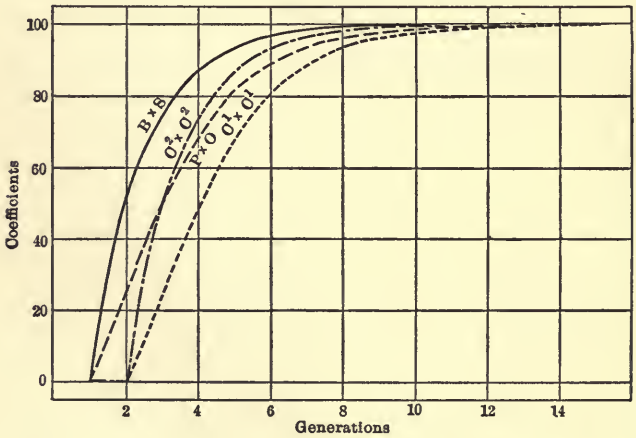



FIG. 1. Curves of inbreeding, showing (*a*) the limiting case of continued brother × sister breeding, wherein the successive coefficients of inbreeding have the maximum values; (*b*) continued parent × offspring mating; (*c*) continued first-cousin × first-cousin mating where the cousinship is double ($C^2 \times C^2$), and (*d*) continued first-cousin × first-cousin mating where the cousinship is single ($C^1 \times C^1$). The continued mating of uncle × niece gives the same curve as $C^1 \times C^1$.

(having an obligate bisexual type of reproduction) that its pedigree on analysis will give $Z_3 > 87.5$. If, therefore, the coefficients of Table I are plotted, the result will be the maximum limiting *curve* of inbreeding. This curve is shown in Fig. 1.

In all actually realized pedigrees except those in which there has been continued brother × sister breeding the curve of inbreeding found will lie wholly or in part below the maximum curve shown in Fig. 1.

### ILLUSTRATION II. PARENT × OFFSPRING BREEDING

The next illustration of the application of coefficients of inbreeding will be the general case of back-crossing, that is, the mating of parent × offspring. Such a case is illustrated in the hypothetical pedigree, Table II.

Here it will be seen that *b*, the dam of *y*, is a daughter of *a*, who is also the sire of *y*, and that in each preceding generation every daughter is bred back to her sire. Proceeding as before to calculate the coefficients of inbreeding, we have, first,

$$Z_0 = \frac{100\ (2-2)}{2} = 0.$$

In forming the expression for $Z_1$ we are met by the fact in determining $q_{n+1}$ for generation 2 that the individual *a* has already appeared once and been counted as a "different" ancestor in

PEDIGREE TABLE II (HYPOTHETICAL)

TO ILLUSTRATE THE BREEDING OF PARENT × OFFSPRING

| | | | | |
|---|---|---|---|---|
| | | | $f$ | $m$ |
| | | | | $n$ |
| | | $d$ | $g$ | $f$ |
| | | | | $l$ |
| | $a$ | | $f$ | |
| | | | $d$ | $g$ |
| | | $e$ | | $d$ |
| | | | $i$ | $o$ |
| | | | $d$ | $f$ |
| | | $a$ | | $g$ |
| | | | $e$ | $d$ |
| | $b$ | | | $i$ |
| | | | $a$ | $d$ |
| | | $c$ | | $e$ |
| | | | $k$ | $a$ |
| | | | | $s$ |
| Generation number . . . | 1 | 2 | 3 | 4 |

generation 1. Therefore it will not be counted a second time in generation 2, and we have

$$Z_1 = \frac{100\,(4-3)}{4} = 25,$$

and by the same process,

$$Z_2 = \frac{100\,(8-4)}{8} = 50,$$

$$Z_3 = \frac{100\,(16-5)}{16} = 68.75,$$

$$Z_4 = \frac{100\,(32-6)}{32} = 81.25,$$

and so forth.

The values of the successive coefficients for parent × offspring for 16 ancestral generations are given in Table 2.

By comparison of this table with Table 1 it is evident that while the increase in intensity of inbreeding is not so rapid in the first few ancestral generations by this parent × offspring type of breeding as with the brother × sister type, by the time the tenth ancestral generation is reached the values are for practical purposes the same.

The curve of inbreeding for continued parent × offspring breeding is shown in Fig. 1.

### ILLUSTRATION III. FIRST-COUSIN × FIRST-COUSIN BREEDING

As a third illustration may be taken the case of continued cousin mating. Such breeding represents the next step in decreasing intensity of inbreeding beyond the parent × offspring type.

TABLE 2

VALUES OF THE SUCCESSIVE COEFFICIENTS OF INBREEDING IN THE CASE OF CONTINUED PARENT × OFFSPRING MATING

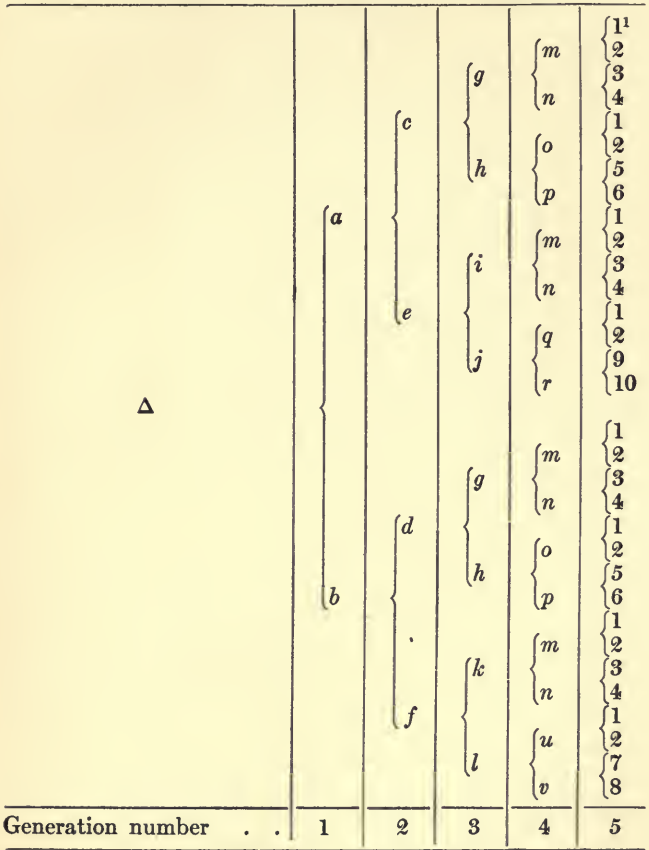
| COEFFICIENT OF INBREEDING | ANCESTRAL GENERATIONS INCLUDED | NUMERICAL VALUE OF COEFFICIENT |
|---|---|---|
| $Z_0$ | 1 | 0 |
| $Z_1$ | 2 | 25 |
| $Z_2$ | 3 | 50 |
| $Z_3$ | 4 | 68.75 |
| $Z_4$ | 5 | 81.25 |
| $Z_5$ | 6 | 89.06 |
| $Z_6$ | 7 | 93.75 |
| $Z_7$ | 8 | 96.48 |
| $Z_8$ | 9 | 98.05 |
| $Z_9$ | 10 | 98.93 |
| $Z_{10}$ | 11 | 99.41 |
| $Z_{11}$ | 12 | 99.68 |
| $Z_{12}$ | 13 | 99.83 |
| $Z_{13}$ | 14 | 99.91 |
| $Z_{14}$ | 15 | 99.95 |
| $Z_{15}$ | 16 | 99.97 |

There are, of course, two possible sorts of first cousins, single and double. In the first case one of the parents of any individual is a brother (or sister) to one of the parents of the other individual in the mating. In the second case both the parents occupy this relation to the parents of the other individual in the mating.

These two sorts of first cousinship are shown in Pedigree Tables III and IV.

## PEDIGREE TABLE III (HYPOTHETICAL)

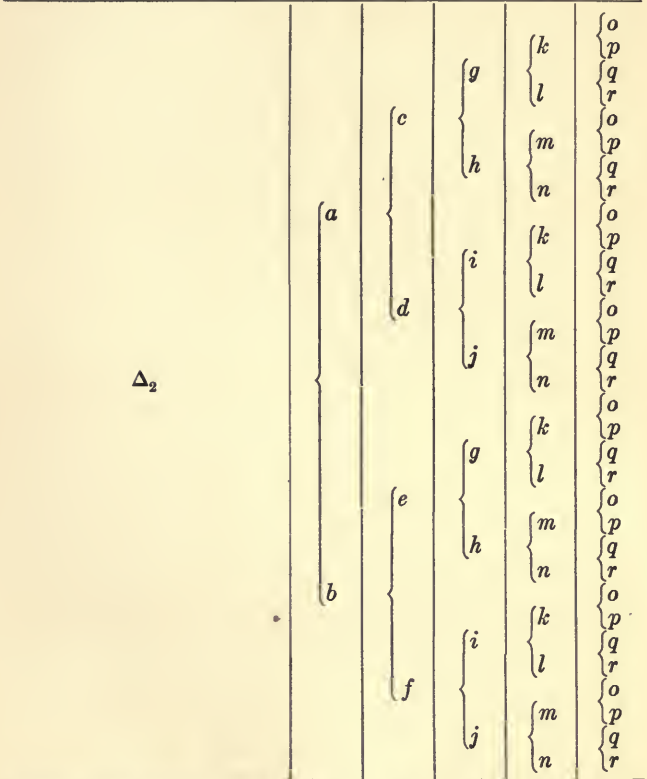
To ILLUSTRATE THE CONTINUED BREEDING OF FIRST-COUSIN × FIRST-COUSIN — SINGLE COUSINS




[1] Owing to the limitation of the alphabet, resort is had to numbers to designate individuals in this generation.

PEDIGREE TABLE IV (HYPOTHETICAL)

To Illustrate the Continued Breeding of First-Cousin × First-Cousin — Double Cousins




| Generation number . . | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|

The values of the successive coefficients of inbreeding for these cases are given in Table 3.

The calculation is carried out in accordance with the same principles as have been illustrated in the previous cases. We have, from Pedigree Table III,

$$Z_0 = \frac{100\ (2—2)}{2} = 0,$$

and

$$Z_1 = \frac{100\ (4—4)}{4} = 0,$$

since in generations 1 and 2 there are two and four *different* ancestors respectively.

$$Z_2 = \frac{100\ (8—6)}{8} = 25,$$

since in generation 3 the two individuals $g$ and $h$ each appear twice, and by our rule any ancestor is only counted once.

$$Z_3 = \frac{100\ (16—8)}{16} = 50,$$

since in generation 4 the individuals $m$ and $n$ appear four times and are only counted as different ancestors once each, and individuals $o$ and $p$ each appear twice.

The data of Table 3 are given graphically in Fig. 1, together with the curve for brother × sister and parent × offspring.

From the table and figure it is seen that with continued inbreeding according to any one of these four types the coefficient approaches the value

TABLE 3

VALUES OF THE SUCCESSIVE COEFFICIENTS OF INBREEDING IN THE CASE OF CONTINUED COUSIN MATING

| COEFFICIENT OF INBREEDING | ANCESTRAL GENERATION INCLUDED | COEFFICIENT FOR SINGLE COUSINS | COEFFICIENT FOR DOUBLE COUSINS |
|---|---|---|---|
| $Z_0$ | 1 | 0 | 0 |
| $Z_1$ | 2 | 0 | 0 |
| $Z_2$ | 3 | 25.00 | 50.00 |
| $Z_3$ | 4 | 50.00 | 75.00 |
| $Z_4$ | 5 | 68.75 | 87.50 |
| $Z_5$ | 6 | 81.25 | 93.75 |
| $Z_6$ | 7 | 89.06 | 96.88 |
| $Z_7$ | 8 | 93.75 | 98.44 |
| $Z_8$ | 9 | 96.48 | 99.22 |
| $Z_9$ | 10 | 98.05 | 99.61 |
| $Z_{10}$ | 11 | 98.93 | 99.80 |
| $Z_{11}$ | 12 | 99.41 | 99.90 |
| $Z_{12}$ | 13 | 99.68 | 99.95 |
| $Z_{13}$ | 14 | 99.83 | 99.98 |
| $Z_{14}$ | 15 | 99.91 | 99.99 |
| $Z_{15}$ | 16 | 99.95 | 99.994 |

100. The rate of approach is different, however, in the different cases. The curves fall into two pairs. The brother × sister and the double cousin curves are precisely alike so far as concerns their curvature or shape at any given point. Similarly, the parent × offspring and single cousin curves are of the same shape. *The essential point of difference is that the cousin curves lag a generation behind the others.*

### ILLUSTRATION IV. UNCLE × NIECE MATING

Let us now consider the question of the degree of inbreeding following continued matings of the avuncular type of relationship. Pedigree Table V gives a pedigree in which each mating is of uncle × niece.

From this table it appears that the values of the coefficients of inbreeding will be exactly the same for this type of mating as in the case of single cousin mating. Or, in other words, $Z$'s form the following series.

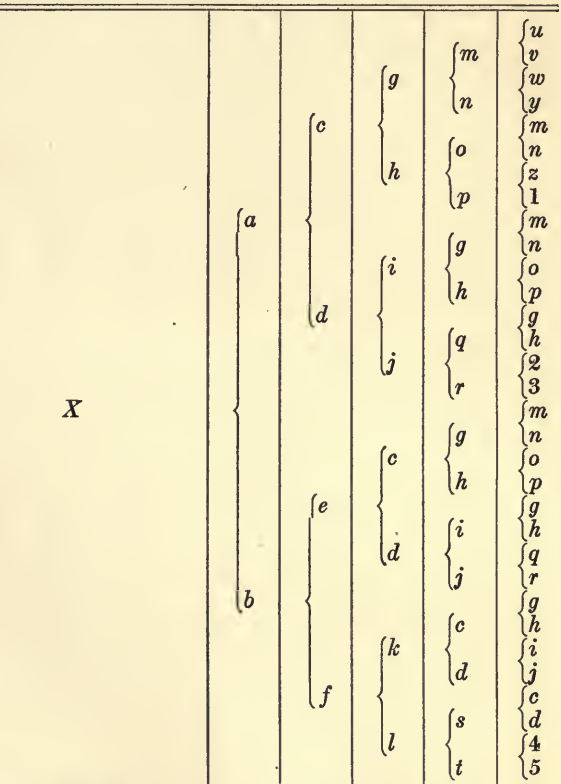
VALUES OF COEFFICIENTS OF INBREEDING FOR CONTINUED UNCLE × NIECE MATING

| COEFFICIENT | NUMBER OF ANCESTRAL GENERATIONS | VALUE OF COEFFICIENT |
|---|---|---|
| $Z_0$ | 1 | 0 |
| $Z_1$ | 2 | 0 |
| $Z_2$ | 3 | 25.00 |
| $Z_3$ | 4 | 50.00 |
| $Z_4$ | 5 | 68.75 |
| $Z_5$ | 6 | 81.25 |
| etc. | etc. | etc. as in Table 3 |

From the data presented it is clear that inbreeding continued for about ten generations, quite regardless of the *type* of mating provided only it be *continuously* followed, leads to within one or two per cent of complete "concentration of

PEDIGREE TABLE V (HYPOTHETICAL)

TO ILLUSTRATE THE MATING OF UNCLE × NIECE




| Generation number . . | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|

blood." The bearing of this result upon the general question of the degree of inbreeding which

exists in the ancestry of our domestic animals to-day is obvious. To consider but a single case: In 1789 [1] a law was passed prohibiting the importation of cattle into the island of Jersey. Hence it follows that all pure-bred Jersey cattle of the present time must be the descendants of the relatively few animals on the island in 1790. Taking three years as about the average generation interval in cattle, this means about forty generations since the island was closed to importation. The concentration of lines of descent which must have occurred in this time merely by the dropping of lines and quite regardless of the type of mating is obvious.

### ILLUSTRATION V. THE PEDIGREE OF THE JERSEY COW, BESS WEAVER (155121)

Leaving now the hypothetical cases, we may consider some pedigrees of actually existing animals. For a first illustration of this sort the Jersey cow, Bess Weaver, may be taken. Her pedigree through four ancestral generations is shown in Pedigree Table VI.

In the twelfth ancestral generation the theoretically possible number of different ancestors is 4096. In a relatively long pedigree, such as arises in dealing with registered cattle, it would obviously be an extremely tedious business to

[1] *Teste* Rees's Encyclopedia and H. S. Redfield, *Natl. Stockman and Farmer*, December 15, 1892.

## PEDIGREE TABLE VI

| Sex ♀ | | | | | |
|---|---|---|---|---|---|
| ♂ | | No. 35913 ♂<br>Sisera's Stoke Pogis | No. 26271 ♂<br>Juno's Stoke Pogis | No. 14207 ♂<br>Alphea's Stoke Pogis | |
| | | | | No. 14436 ♀<br>Carlo's Juno | |
| | | | No. 37346 ♀<br>Sisera | No. 1881 ♂<br>Duchess Stoke Pogis | |
| | | | | No. 6246 ♀<br>Edith Darby | |
| | Davy Stoke Pogis | No. 79860 ♀<br>Baltimore | No. 19350 ♂<br>Patrick Fawkes | No. 10469 ♂<br>Regal Koffee | |
| | | | | No. 21574 ♀<br>Kermesse | |
| | | | No. 17900 ♀<br>Avoca 2nd | No. 3286 ♂<br>Champion's Son | |
| BESS WEAVER | No. 53598 | | | No. 17769 ♀<br>Avoca | |
| ♀ | | No. 35913 ♂<br>⊗<br>Sisera's Stoke Pogis | No. 26271 ♂<br>×<br>Juno's Stoke Pogis | No. 14207 ♂<br>×<br>Alphea's Stoke Pogis | |
| | | | | No. 14436 ♀<br>× Carlo's Juno | |
| | | | No. 87346 ♀<br>×<br>Sisera | No. 18811 ♂<br>×<br>Duchess Stoke Pogis | |
| | | | | No. 6246 ♀<br>× Edith Darby | |
| | Peg Weaver | No. 126626 ♀<br>Kate Weaver | No. 36382 ♂<br>General Kelly | No. 19350 ♂<br>⊗ Patrick Fawkes | |
| | | | | No. 95606 ♀<br>⊗ Balm | |
| No. 155121 | No. 126629 | | No. 95606 ♀<br>Balm | No. 7056 ♂<br>America's Champion | |
| | | | | No. 95605 ♀<br>Maid of Gilead 2nd | |

determine the value of $q$ by direct counting, as has been done in the preceding simpler illustrations. The calculation of the coefficients of inbreeding may be greatly simplified in the case of long pedigrees by a system of counting which makes the *line of descent* the unit rather than the individual. This system is used in the above pedigree as an illustration of method, although only 4 ancestral generations are here considered. While each individual animal which is eliminated because of previous appearances in a lower ancestral generation is marked with an ×, those at the apex of a line of descent are marked with a cross within a circle. These latter are all that need to be counted directly. Thus the bull Sisera's Stoke Pogis first appears in the second ancestral generation as the sire of Davy Stoke Pogis. He next appears (here marked with a cross within a circle) in the same generation as the sire of Peg Weaver. He will, by the general rule of coefficients of inbreeding, not be counted as a "different" ancestor the second time in this generation. But this automatically eliminates his two parents in the third ancestral generation, his four grandparents in the fourth generation, and so on until in the twelfth generation 1024 ancestors of Sisera's Stoke Pogis will be so eliminated. The same consideration applies in every other like case.

Practically, then, the method of dealing with a pedigree of this sort is first to go through and

indicate in a distinctive way every primary [1] reappearance of individuals. Then form a table on the plan of Table 4, the character of which is so obvious as not to need detailed explanation.

This table is to be read in the following way: Because of the reappearance of Sisera's Stoke Pogis in the 2d ancestral generation Bess Weaver has 1 fewer ancestors in that generation than she would have had in the entire absence of inbreeding; 2 fewer in the 3d generation, and so on. The totals of the columns of this table are the values, for each generation, of

$$p_{n+1} - q_{n+1}.$$

in (*i*). These totals, multiplied by 100, have then merely to be divided by $p_{n+1}$ in order to obtain the successive $Z$'s. The whole operation may be very quickly carried out. It is not necessary, in fact, to fill out the whole of the later columns of the table; the entries may be cumulated.

For the present pedigree we have

$$Z_0 = 0, \text{ as always.}^2$$

$$Z_1 = \frac{100(1)}{4} = 25 \text{ per cent,}$$

[1] By "primary" reappearance in the pedigree is meant a reappearance as the sire or dam of an individual which has not itself appeared before in the lower ancestral generations. Thus Patrick Fawkes makes a *primary* reappearance in the fourth ancestral generation as the sire of General Kelly, a bull which is not found in any generation below the third.

[2] The apparent paradox implied in the fact that $Z_0$ must always

TABLE 4

WORKING TABLE USED IN CALCULATING THE COEFFICIENTS OF INBREEDING FOR PEDIGREE TABLE VI

| ANIMAL | ANCESTRAL GENERATION. | | |
|---|---|---|---|
| | 2 | 3 | 4 |
| Sisera's Stoke Pogis . . . | 1 | 2 | 4 |
| Patrick Fawkes . . . . | — | — | 1 |
| Balm . . . . . | — | — | 1 |
| Totals . . . . . | 1 | 2 | 6 |

$$Z_2 = \frac{100(2)}{8} = 25 \text{ per cent,}$$

$$Z_3 = \frac{100(6)}{16} = 37.5 \text{ per cent.}$$

From these values it is seen that in the first four ancestral generations the cow Bess Weaver is 37.5 per cent inbred. This is a perfectly definite figure, directly comparable with similar constants for other animals. Of course, if we were to go back more generations we should find Bess Weaver still more inbred, that is, the coefficients would

be zero, or in other words that in the first ancestral generation, considered *alone*, there is no inbreeding will be cleared up, if it strikes the reader as paradoxical, by a reconsideration of the general principle numbered 5 on p. 110. The point of course is that it is impossible to say whether the parents are or are not related to one another until something is known of *their* parentage, or in other words, until a *second* ancestral generation is considered.

grow larger with each case of the mating of relatives. Since the case is cited here merely for illustration of method, four generations only are considered.

## II. Inbreeding and Relationship Coefficients

In the discussion which has preceded no mention has been made of an important consideration which arises in connection with the analysis of inbreeding by means of the coefficients described. This further problem, to which we may now turn, may be stated in the following way:

The pedigree of an individual consists of two halves. One of these halves is made up of the sire and his ancestors; the other of the dam and her ancestors. Following the conception of inbreeding set forth in detail in the earlier papers of this series, it is plain that the values of the coefficients of inbreeding for a particular pedigree are composed of the following elements:

1. The occurrence of the same individual animals more than once on the sire's side of the pedigree only.

2. The occurrence of the same individual animals more than once on the dam's side of the pedigree only.

3. The reappearance of animals which appear

first on one side of the pedigree (either the sire's or the dam's) then on the other side.

If only 1 and 2 are to be found in the pedigree, it means that the sire and the dam are totally unrelated (within the limits covered by the pedigree in the particular case). On the other hand, the occurrence of 3 means that sire and dam are in some degree related, and that a portion of the observed inbreeding arises because of that fact. Now the coefficients of inbreeding, in and of themselves, tell nothing about what proportionate part has been played by these three elements in reaching the final result. It is a matter of great importance to have information on this point, because of its genetic significance. It is now proposed to describe a general method for obtaining this desired information.

The first step in the method, stated briefly, is to break up the pedigree elimination table formed to get the successive values of $p_{n+1}$, in our former notation, into four different parts. One of these parts will include the primary reappearance on the sire's side of the pedigree of such animals as appear first on the same side. This may be called the "male only" table. The second part will include the primary reappearance on the dam's side of such animals as first appear on the same side. This is the "female only" table. The third part will include the primary reappearance on the dam's side of such animals as first

appear on the sire's side. The fourth part is the reverse of the third. These last two may be called the "cross tables." The sums of the totals of these partial tables will give the total $p_{n+1} - q_{n+1}$ values for the successive generations.

The formation of the tables on this plan may be illustrated with some examples. These examples will also show the skeleton method of writing pedigree elimination tables, which saves much labor. It consists simply in doubling the total of the column for each generation rather than the separate items.

The pedigree for 12 ancestral generations of the Jersey bull King Melia Rioter 14th (103901) may be taken as the first illustration.

From these tables it is obvious that a very considerable portion of the inbreeding shown in the pedigree of King Melia Rioter 14th arises from the fact that his sire and dam were closely related. Furthermore, both sire and dam are closely inbred in their own lines. The curve of total inbreeding in this case is shown in Fig. 2, along with the curves for continued brother × sister, and parent by offspring.

Table 7 is clearly the one which demands special attention. As will shortly appear, it is the most important for the theory of inbreeding. Let us attempt its analysis. Just what does the first entry mean genetically? It states that King Melia Rioter, an animal which first appeared on

## TABLE 5

PARTIAL PEDIGREE ELIMINATION TABLE FOR KING MELIA RIOTER 14TH, SHOWING THE PRIMARY REAPPEARANCES ON THE SIRE'S SIDE OF THE PEDIGREE OF ANIMALS WHICH FIRST APPEAR ON THAT SIDE

| GENERATION | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Melia Ann's Son | | | 1 | (2)[1] | | | | | | | |
| Melia Ann 3d | | | | 1 | (6)[1] | | | | | | |
| Lucy's Stoke Pogis | | | | | 3 | | | | | | |
| Melia Ann | | | | | 2 | | | | | | |
| St. Lambert Boy | | | | | 1 | | | | | | |
| Letty Rioter | | | | | 1 | | | | | | |
| Allie of St. Lambert | | | | | 1 | | | | | | |
| Lord Aylmer | | | | | 1 | | | | | | |
| Amelia 2d | | | | | 1 | (32)[1] | | | | | |
| Victor Hugo | | | | | | 1 | | | | | |
| Oakland's Nora | | | | | | 1 | | | | | |
| Stoke Pogis 3d | | | | | | 1 | | | | | |
| Bachelor of St Lambert | | | | | | 1 | | | | | |
| Sir George of St. Lambert | | | | | | 1 | | | | | |
| Diana's Rioter | | | | | | 1 | | | | | |
| Orloff | | | | | | 1 | | | | | |
| Lorne | | | | | | 1 | | | | | |
| Hugo's Victoria | | | | | | 1 | (82)[1] | | | | |
| Victor Hugo | | | | | | | 1 | | | | |
| Pauline | | | | | | | 1 | | | | |
| Canada's John Bull | | | | | | | 1 | | | | |
| Oakland's Nora | | | | | | | 1 | | | | |
| Stoke Pogis 3d | | | | | | | 7 | | | | |
| Kathleen of St. Lambert | | | | | | | 1 | | | | |
| Lord Lisgar | | | | | | | 4 | | | | |
| Lucy of St. Lambert | | | | | | | 2 | | | | |
| Diana of St. Lambert | | | | | | | 1 | | | | |
| Pet of St. Lambert | | | | | | | 1 | | | | |
| Orloff | | | | | | | 1 | | | | |
| Bachelor of St. Lambert | | | | | | | 1 | | | | |

[1] In this and the following tables the numbers in brackets are in each case twice the sum of the numbers in the preceding column. They represent the accumulated ancestral reduplication up to the generation in question.

TABLE 5 — *Continued*

PARTIAL PEDIGREE ELIMINATION TABLE FOR KING MELIA RIOTER 14TH, SHOWING THE PRIMARY REAPPEARANCES ON THE SIRE'S SIDE OF THE PEDIGREE OF ANIMALS WHICH FIRST APPEAR ON THAT SIDE

| GENERATION | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Ida of St. Lambert | | | | | | | 1 | (210)[1] | | | |
| Victor Hugo . . . | | | | | | | | 2 | | | |
| Stoke Pogis 3d . . | | | | | | | | 2 | | | |
| Lord Lisgar . . . | | | | | | | | 3 | | | |
| Lorne . . . . . | | | | | | | | 1 | | | |
| Amelia . . . . | | | | | | | | 1 | (438)[1] | | |
| Lord Lisgar . . . | | | | | | | | | 1 | | |
| Pride of Windsor . | | | | | | | | | 2 | | |
| Laval . . . . . | | | | | | | | | 1 | | |
| Amelia . . . . | | | | | | | | | 2 | | |
| Victor Hugo . . | | | | | | | | | 3 | (894)[1] | |
| Laval . . . . . | | | | | | | | | | 1 | |
| Amelia . . . . | | | | | | | | | | 1 | |
| Lisette . . . . | | | | | | | | | | 1 | |
| Berthe . . . . | | | | | | | | | | 1 | |
| Totals . . . | | | 1 | 3 | 16 | 41 | 105 | 219 | 447 | 898 | 1,796 |

TABLE 6

PARTIAL PEDIGREE ELIMINATION TABLE FOR KING MELIA RIOTER 14TH, SHOWING THE PRIMARY REAPPEARANCES ON THE DAM'S SIDE OF THE PEDIGREE OF ANIMALS WHICH FIRST APPEAR ON THAT SIDE

| GENERATION | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| King's Rioter Lad | | | | 1 | 2 | 4 | 8 | 16 | 32 | 63 | 128 |

[1] In this and the following tables the numbers in brackets are in each case twice the sum of the numbers in the preceding column. They represent the accumulated ancestral reduplication up to the generation in question.

the sire's side of the pedigree, reappeared in the second ancestral generation on the dam's side. What this clearly means is that at least one half of all the dam's ancestors, in the third and higher ancestral generations, *are identically the same animals as are ancestors of the sire.* The next

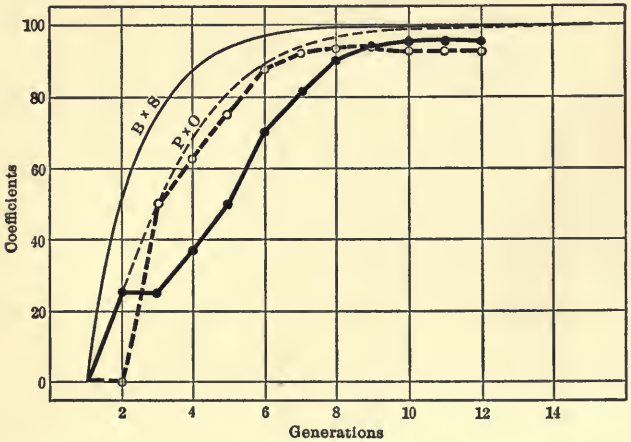



FIG. 2.— Diagram showing (*a*) the total inbreeding (heavy solid line) and (*b*) the relationship (heavy broken line) curves for the Jersey Bull, King Melia Rioter 14th. The high order of the inbreeding and relationship between the sire and dam in this case is evident by comparison with the lighter lines, which give the maximum values for continued brother × sister and parent × offspring breeding.

entry in Table 7 indicates that in the fourth and higher ancestral generations at least $\frac{5}{8}$ of all the dam's ancestors were the same individual animals as were also ancestors of the sire. One half of them were the same before the reappearance of

St. Lambert's Rioter King. He makes up the additional $\frac{1}{8}$ of the dam's ancestry.

## TABLE 7

PARTIAL PEDIGREE ELIMINATION TABLE FOR KING MELIA RIOTER 14TH, SHOWING THE PRIMARY REAPPEARANCES ON THE DAM'S SIDE OF THE PEDIGREE OF ANIMALS WHICH FIRST APPEAR ON THE SIRE'S SIDE.

| GENERATION | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| King Melia Rioter . | 1 | (2) | (4) | | | | | | | | |
| St. Lambert's Rioter King . . . . . | | | 1 | (10) | | | | | | | |
| King of St. Lambert | | | | 1 | | | | | | | |
| St. Lambert Boy . | | | | 1 | (24) | | | | | | |
| St. Lambert Boy . | | | | | 2 | | | | | | |
| Oakland's Nora . . | | | | | 1 | | | | | | |
| St. Lambert's Rioter King . . . . . | | | | | 1 | (56) | | | | | |
| St. Lambert Boy . | | | | | | 1 | | | | | |
| King of St. Lambert | | | | | | 1 | | | | | |
| St. Lambert's Letty | | | | | | 1 | (118) | | | | |
| Letty Coles 2d . . | | | | | | | 1 | (238) | | | |
| King of St. Lambert | | | | | | | | 1 | | | |
| Louise's Grace . . | | | | | | | | 1 | | | |
| Totals . . . | 1 | 2 | 5 | 12 | 28 | 59 | 119 | 240 | 480 | 960 | 1,920 |

## TABLE 8

SUMMARIZED PEDIGREE ELIMINATION TABLE FOR KING MELIA RIOTER 14TH

| GENERATION | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ♂ only . . | | | 1 | 3 | 16 | 41 | 105 | 219 | 447 | 898 | 1,796 |
| ♀ only . . | | | | 1 | 2 | 4 | 8 | 16 | 32 | 64 | 128 |
| Cross . . . | 1 | 2 | 5 | 12 | 28 | 59 | 119 | 240 | 480 | 960 | 1,920 |
| Together . | 1 | 2 | 6 | 16 | 46 | 104 | 232 | 475 | 959 | 1,922 | 3,844 |

From this we have, for the inbreeding coefficients,

$$
\begin{aligned}
Z_0 &= 0 \\
Z_1 &= 25.00 \\
Z_2 &= 25.00 \\
Z_3 &= 37.50 \\
Z_4 &= 50.00 \\
Z_5 &= 71.88 \\
Z_6 &= 81.25 \\
Z_7 &= 90.63 \\
Z_8 &= 92.77 \\
Z_9 &= 93.65 \\
Z_{10} &= 93.85 \\
Z_{11} &= 93.85
\end{aligned}
$$

These facts will possibly be made clearer to those not actually working much with pedigrees by Table VII, which gives the first four ancestral generations [1] of the pedigree of King Melia Rioter 14th.

Generalizing the above reasoning we get the following result:

In $A_3$, and higher ancestral generations, $\frac{2}{4} = 50.00$ per cent of the dam's ancestors are animals which are also ancestors of the sire.

[1] In the study of pedigrees stress is naturally laid on the ancestral generations, rather than on the filial, as in breeding experiments. It becomes very convenient to have a brief designation for ancestral generations, in the same way that $F_1$, $F_2$, etc., are used to denote filial generations. I would suggest the use of the letter $A$ with subnumbers for this purpose. We then have $A_1$ denoting the parental generation, $A_2$ the grandparental, $A_3$ the great-grand-parental, etc.

In $A_4$, and higher ancestral generations, $\frac{5}{8}$ = 62.50 per cent of the dam's ancestors are animals which are also ancestors of the sire.

In $A_5$, and higher ancestral generations, $\frac{12}{16}$ = 75.00 per cent of the dam's ancestors are animals which are also ancestors of the sire.

In $A_6$, and higher ancestral generations, $\frac{28}{32}$ = 87.50 per cent of the dam's ancestors are animals which are also ancestors of the sire.

In $A_7$, and higher ancestral generations, $\frac{59}{64}$ = 92.19 per cent of the dam's ancestors are animals which are also ancestors of the sire.

In $A_8$, and higher ancestral generations, $\frac{119}{128}$ = 92.97 per cent of the dam's ancestors are animals which are also ancestors of the sire.

In $A_9$, and higher ancestral generations, $\frac{240}{256}$ = 93.75 per cent of the dam's ancestors are animals which are also ancestors of the sire.

In $A_{10}$, and higher ancestral generations, 93.75 per cent of the dam's ancestors are animals which are also ancestors of the sire.

In $A_{11}$, and higher ancestral generations, 93.75 per cent of the dam's ancestors are animals which are also ancestors of the sire.

In $A_{12}$, and higher ancestral generations, 93.75 per cent of the dam's ancestors are animals which are also ancestors of the sire.

These percentages are quantities of a good deal of interest. They measure the degree in which King Melia Rioter 14th's sire and dam were

## PEDIGREE TABLE VII

PEDIGREE FOR FOUR ANCESTRAL GENERATIONS OF KING MELIA RIOTER 14TH

| SEX | | | | | |
|---|---|---|---|---|---|
| ♂ | | No. 63200 ♂ Marjorie Melia Ann's Son | No. 56581 ♂ Melia Ann's King | No. 22041 ♂ Melia Ann's Son | |
| | | | | No. 100775 ♀ Lottie Melia Ann | |
| | | | No. 157263 ♀ Marjorie Melia Ann | No. 22041 ♂ ⊗ Melia Ann's Son | |
| | | | | No. 905883 ♀ Mary Melia Ann | |
| | KING MELIA RIOTER | No. 181544 ♀ Letty Silver Hair | No. 58169 ♂ King of All Kings | No. 54896 ♂ St. Lambert's Rioter King | |
| | | | | No. 114804 ♀ St. Lambert's Letty | |
| | | | No. 148456 ♀ Exile's Silver Hair | No. 32559 ♂ Exile of St. Anne's | |
| | | | | No. 60449 ♀ Silver Hair 4th | |
| ♂ | No. 73104 | No. 73104 ♂ ⊗ King Melia Rioter | No. 63200 ♂ × Marjorie Melia Ann's Son | No. 56581 ♂ × Melia Ann's King | |
| | | | | No. 157263 ♀ × Marjorie Melia Ann | |
| | | | No. 181544 ♀ × Letty Silver Hair | No. 58169 ♂ × King of All Kings | |
| | | | | No. 148456 ♀ × Exile's Silver Hair | |
| KING MELIA RIOTER 14th | | No. 219360 ♀ Dula Riotress Maid | No. 62098 ♂ King Rioter's Lad | No. 54896 ♂ ⊗ St. Lambert's Rioter King | |
| | | | | No. 142296 ♀ King's Riotress Nora | |
| | | | No. 218796 ♀ St. Lambert's Dula Riotress | No. 57778 ♂ St. Lambert's Boy | |
| | | | | No. 174761 ♀ Rioter Lad's First Daughter | |
| No. 103901 | No. 262996 | LARISSA MARIE | | | |

related to each other. Community of ancestry is the basis of kinship.

Percentages derived in the way shown above, from cross pedigree elimination tables, I have called *coefficients of relationship*, and designated by the letter $K$, with appropriate sub-numbers referring to the generation. These relationship coefficients are, with some limitations, independent of the inbreeding coefficients in the values they may take, though the two will usually be correlated to some degree. It is, however, possible to have a high value of $Z$ with $K = 0$.

The most important feature of the relationship coefficients is found in their genetic implications. This can be indicated best by an illustration. Let us consider the case of the maximum possible degree of inbreeding with $K = 0$. This will be found when the sire and the dam are each inbred to the highest possible degree (continued brother $\times$ sister mating), but are in no way related to each other. Such a case would be afforded if a Jersey bull, the product of continued brother $\times$ sister mating, was bred to a Holstein cow, which was also the product of a continued brother $\times$ sister breeding. Clearly $K$ would be 0, since no animal on one half of the pedigree could ever appear on the other. The values of the successive coefficients of inbreeding ($Z$'s) in such a case are shown in Table 9, where they are compared with the coefficients of inbreeding in com-

TABLE 9

COMPARING THE MAXIMUM POSSIBLE VALUES OF THE COEFFICIENTS OF INBREEDING ($Z$) WHEN THE COEFFICIENT OF RELATIONSHIP $K$ EQUALS ($a$) ZERO AND ($b$) 100

| GENERATION | MAXIMUM POSSIBLE VALUE OF $Z$ WHEN $K = 0$ | MAXIMUM POSSIBLE VALUE OF $Z$ WHEN $K = 100$ |
|---|---|---|
| $A_1$ | 0 | 0 |
| $A_2$ | 0 | 50.00 |
| $A_3$ | 50.00 | 75.00 |
| $A_4$ | 75.00 | 87.50 |
| $A_5$ | 87.50 | 93.75 |
| $A_6$ | 93.75 | 96.88 |
| $A_7$ | 96.88 | 98.44 |
| $A_8$ | 98.44 | 99.22 |
| $A_9$ | 99.22 | 99.61 |
| $A_{10}$ | 99.61 | 99.80 |

plete continued brother × sister mating, where $K = 100$.[1]

From this it appears that an individual may be inbred in 10 generations to within two tenths of one per cent as intensely, measured by the coefficients of inbreeding, if his sire and dam are in no way related, as he would be if his sire and dam were brother and sister. But clearly the germinal constitution of the individual produced would be quite different in the two calves. This point is

[1] Since, of course, *all* of a sister's ancestors are identical with her brother's.

so evident as to need no elaboration. It has been brought out by East and Hayes.[1]

The values of the $K$'s for a particular pedigree evidently furnish a rough index of the probability that the two germ plasms which unite to form an individual are alike in their constitution. This will follow because of the fact that the probability of likeness of germinal constitution in two individuals must tend to increase as the number of ancestors common to the two increases. Just what is the law of this increase in probability is a problem in Mendelian mathematics which has not yet been worked out. The general fact, however, seems quite sure.

From the above discussion it seems plain that in reaching a numerical measure of the degree of inbreeding it is not sufficient to consider coefficients of inbreeding alone. The coefficients of relationship must also be taken into account.

It is suggested that the two constants be written together for each generation, the coefficient of inbreeding being followed by the coefficient of relationship in brackets.

Thus we have

INBREEDING AND RELATIONSHIP COEFFICIENTS OF KING MELIA RIOTER 14TH

$$Z_0\ (K_1) = 0 \qquad (0)$$
$$Z_1\ (K_2) = 25 \qquad (0)$$

[1] U. S. Dept. Agr. Bur. Plant Industry, Bul. 243, 1912.

$$Z_2\ (K_3) = 25.00\ (50.00)$$
$$Z_3\ (K_4) = 37.50\ (62.50)$$
$$Z_4\ (K_5) = 50.00\ (75.00)$$
$$Z_5\ (K_6) = 71.88\ (87.50)$$
$$Z_6\ (K_7) = 81.25\ (92.19)$$
$$Z_7\ (K_8) = 90.63\ (92.97)$$
$$Z_8\ (K_9) = 92.77\ (93.75)$$
$$Z_9\ (K_{10}) = 93.65\ (93.75)$$
$$Z_{10}\ (K_{11}) = 93.85\ (93.75)$$
$$Z_{11}\ (K_{12}) = 93.85\ (93.75)$$

The physical meaning of these expressions is simple and straightforward. $Z_4$ ($K_5$) tells that in the 5th ancestral generation of King Melia Rioter 14th he had only one half as many different ancestors as was possible for that generation, and of his ancestors three fourths were common to his sire and his dam. However one looks at the matter there can be no denial that King Melia Rioter 14th is a closely inbred animal.

## III. Genetic Bearing

Up to this point the discussion of inbreeding has confined itself entirely to the logical aspects of inbreeding, considered as a mode of mating individuals with relation to their kinship. Nothing has been said, either about the relation, if any, of inbreeding and relationship coefficients to the zygotic or gametic constitution (in the Mendelian sense) of the individual, or about the physiological

effects of inbreeding upon the progeny. This restriction has been deliberate. The primary purpose of what has preceded is to make some contribution to the methodology of the study of the important problem of inbreeding. It has seemed desirable to emphasize the fact that in the investigation of the problem of inbreeding in general there are three elements, which from the standpoint of the logic of the case, are totally distinct and separate. These are:

1. The logical and mathematical characteristics of a system of mating of organisms such that the individual has fewer different ancestors than it would have had under the operation of any other system of mating. Having proper regard for the meaning of words, such a *system of mating*, and that alone, can logically be called *inbreeding*.

2. The necessary *consequences* in respect of the Mendelian constitution of the individual which must follow the continued operation of systems of mating which are inbreeding.

3. The physiological *effect* on the individual consequent upon its having been produced through the operation of a system of mating which is inbreeding.

The first of these three phases of the problem is the one attacked in the preceding sections of this paper. In these the attempt has been made to show in the clearest way of which the writer is capable that inbreeding is a mode or system of

mating organisms, and is, properly speaking, nothing else. Furthermore, precise methods of measuring, and hence of comparing, systems of mating which are inbreeding have been suggested in these preceding sections of the paper.

The other two phases or elements of the general problem are logically consequences or effects of the first. The second phase demands for its solution mathematical analysis, falling in the field of probability, with, of course, the limitations implied by general Mendelian principles. The third phase demands experimental investigation.

I have been at considerable pains to endeavor to make entirely clear and sharply defined the logical elements of the problem of inbreeding for the reason that the distinctions between them seem not always to have been kept in mind.

The investigation of the second phase of the problem has barely been begun. It has been shown by East and Hayes [1] and Jennings [2] that the proportion of homozygotes increases steadily with continued self-fertilization, which represents the absolutely closest possible form of inbreeding. Jennings summarizes his mathematical results

[1] East, E. M., and Hayes, H. K. "Heterozygosis in Evolution and in Plant Breeding." U. S. Dept. Agr. Bur. Plant Ind. Bulletin No. 243, pp. 1–58, 1912.

[2] Jennings, H. S. "Production of Pure Homozygotic Organisms from Heterozygotes by Self-Fertilization." *Amer. Nat.*, Vol. XLVI, pp. 487–491, 1912.

in very convenient form as follows (*loc. cit.*, p. 490):

"Let $x$ = the proportional number of organisms that are pure homozygotes (with respect to all the characters considered),

$y$ = proportion that are heterozygotic with respect to all the characters considered,

$z$ = the proportion that are mixed,

$v$ = the proportion that have any heterozygotic characters.

Then, if $n$ = the number of successive self-fertilizations

and $m$ = the number of pairs of characters,

$$x = \left(\frac{2^n - 1}{2^n}\right)^m, \quad (1)$$

$$y = \left(\tfrac{1}{2}\right)^{mn}, \quad (2)$$

$$z = 1 - (x + y), \quad (3)$$

$$v = 1 - x = \frac{2^{mn} - (2^n - 1)^m}{2^{mn}}. \quad (4)$$"

With continued brother × sister mating the present writer [1] has shown that the proportion of homozygotes also increases, in the manner shown in the following table.

[1] Pearl, R. "On the Results of Inbreeding a Mendelian Population: A Correction and Extension of Previous Conclusions." *Amer. Nat.*, Vol. XLVIII, pp. 57–62, 1914.

Since this was written a paper by Fish announcing an independent working out of the same point has appeared (*Amer. Nat.*, 1915).

Starting, by hypothesis, with all individuals of the population heterozygous, there will be

| After the Indicated Numbers of Generations of Continued Brother × Sister Mating, | The Following Percentage of Homozygotes. |
|---|---|
| 1 | 50.00 |
| 2 | 50.00 |
| 3 | 62.50 |
| 4 | 68.75 |
| 5 | 75.0 |
| 6 | 79.69 |
| 7 | 83.59 |
| 8 | 86.72 |
| 9 | 89.26 |
| etc. | |

These results may be put in the form of a general formula, by means of which the constitution of any generation may be written down from a knowledge of the preceding generation; that is, from a knowledge of the $n-1$th generation the $n$th generation may be at once written down.

This general formula may be developed as follows. A single character pair will be considered, $A$ denoting the dominant character and $a$ the recessive. Equal fertility for all matings is assumed, the number of individuals per family being taken as $2s$, of which $s$ are males and $s$ are females. One family will then make $s$ matings and produce $s$ families in the next generation. Each mating is, by hypothesis, of a brother with his sister.

Starting as before with a pair from a population in which all individuals are of constitution $Aa$, there

will be in the next generation one family of the $AA + Aa + aA + aa$ type. In all succeeding generations there will be six types of families, viz.:

(1) $AA$ families.
(2) $AA + Aa$ families.
(3) $Aa$ families.
(4) $AA + 2\,Aa + aa$ families.
(5) $Aa + aa$ families.
(6) $aa$ families.

The proportionate number of each of these types of families will change in successive generations according to the following system:

Let $o_{n-1}$ denote the number of $AA$ families in the $n - 1$th generation, and

$p_{n-1}$ denote the number of $AA + Aa$ families in the $n - 1$th generation, and

$q_{n-1}$ denote the number of $Aa$ families in the $n - 1$th generation, and

$r_{n-1}$ denote the number of $AA + 2\,Aa + aa$ families in the $n - 1$th generation, and

$u_{n-1}$ denote the number of $Aa + aa$ families, and

$v_{n-1}$ denote the number of $aa$ families.

It will be possible to write down $u$ and $v$ in any case without calculation because of the symmetrical relations of a Mendelian population, since

always under normal conditions such as are assumed in the general treatment, we have

$$\begin{aligned} u_{n-1} &= p_{n-1}, \\ v_{n-1} &= o_{n-1}, \\ u_n &= p_n, \\ v_n &= o_n. \end{aligned}$$

It is necessary, therefore, to consider only the coefficients for the first four types of family. In the $n$th generation the constitution of the population in respect of *families* (not individuals) will be as follows:

Families in the $n$th generation

$$\begin{aligned} = {} & s(o_{n-1} + \tfrac{1}{4}\, p_{n-1} + \tfrac{1}{16}\, r_{n-1})\ AA \text{ families} \\ & + s\ (\tfrac{1}{2}\, p_{n-1} + \tfrac{1}{4}\, r_{n-1})\ AA + Aa \text{ families} \\ & + s\ (\tfrac{1}{8}\, r_{n-1})\ Aa \text{ families} \\ & + s\ (\tfrac{1}{2}\, p_{n-1} + q_{n-1} + \tfrac{1}{4}\, r_{n-1})\ AA \\ & \qquad + 2\ Aa + aa \text{ families} \\ & + s\ (u_n)\ Aa + aa \text{ families} \\ & + s\ (v_n)\ aa \text{ families}. \end{aligned}$$

Or, taking coefficients alone, we have

$$\begin{aligned} o_n &= o_{n-1} + \tfrac{1}{4}\, p_{n-1} + \tfrac{1}{16}\, r_{n-1}, \\ p_n &= \tfrac{1}{2}\, p_{n-1} + \tfrac{1}{4}\, r_{n-1}, \\ q_n &= \tfrac{1}{8}\, r_{n-1}, \\ r_n &= \tfrac{1}{2}\, p_{n-1} + q_{n-1} + \tfrac{1}{4}\, r_{n-1}, \\ u_n &= \tfrac{1}{2}\, u_{n-1} + \tfrac{1}{4}\, r_{n-1} = p_n, \\ v_n &= v_{n-1} + \tfrac{1}{4}\, u_{n-1} + \tfrac{1}{16}\, r_{n-1} = o_n. \end{aligned}$$

Let us see how this formula works out in a con-

crete case. Assume that $2s = 32$, or $s = 16$. Start with a single $AA + 2\,Aa + aa$ family.

Then

$$o_{n-1} = 0,$$
$$p_{n-1} = 0,$$
$$q_{n-1} = 0,$$
$$r_{n-1} = 1.$$

Then in the next generation we shall have

$16\,\{0 + \frac{1}{4}\,(0) + \frac{1}{16}\,(1)\} = 1\ AA$ family
$+ 16\,\{\frac{1}{2}\,(0) + \frac{1}{4}\,(1)\} = 4\ AA + Aa$ families
$+ 16\,\{\frac{1}{8}\,(1)\} = 2\ Aa$ families
$+ 16\,\{\frac{1}{2}\,(0) + 0 + \frac{1}{4}\,(1)\} = 4\,(AA + 2\,Aa + aa)$ families
$+ 4\,(Aa + aa)$ families
$+ 1\ aa$ family.

This is the fact.

In the next generation we shall have

$16\,\{1 + 1 + \frac{1}{16}\,(4)\} = 36\ AA$ families
$+ 16\,\{\frac{1}{2}\,(4) + \frac{1}{4}\,(4)\} = 48\ (AA + Aa)$ families
$+ 16\,\{\frac{1}{8}\,(4)\} = 8\ Aa$ families
$+ 16\,\{\frac{1}{2}\,(4) + 2 + \frac{1}{4}\,(4,\} = 80\,(AA + 2\,Aa + aa)$ families
$+ 48\,(Aa + aa)$ families
$+ 36\,(aa)$ families.

This is the fact.

In the next generation we shall have

$16\,\{36 + \frac{1}{4}\,(48) + \frac{1}{16}\,(80)\} = 16 \times 53 = 848\ AA$ families

$+ 16 \{\frac{1}{2}(48) + \frac{1}{4}(80)\} = 16 \times 44 = 704$ ($AA +$ $Aa$) families

$+ 16 \{\frac{1}{8}(80)\} = 160\ Aa$ families

$+ 16 \{\frac{1}{2}(48) + 8 + \frac{1}{4}(80)\} = 16 \times 52 = 832$ ($AA$ $+ 2\ Aa + aa$) families

$+ 704\ (Aa + aa)$ families

$+ 848\ aa$ families.

Succeeding generations follow the same law and need not be worked out in detail.

So far the discussion has confined itself to families, as this must be the basic unit in the theory of any form of inbreeding. Turning to individuals, we have the following simple relations to pass to individuals:

In the $n$th generation the number of

$AA$ (or $aa$) individuals $= 2\,s\ (o_n) + s\ (p_n) +$ $\frac{1}{2}\,s\ (r_n)$.

$Aa$ (or $aA$) individuals $= 2\,s\ (q_n) + s\ (\frac{1}{2}p_n) +$ $\frac{1}{2}\,s\ (r_n)$.

The first of the above expressions multiplied by 2 gives the total heterozygotes.

The results under certain conditions of brother $\times$ sister and cousin mating have been discussed by Jacobs.[1]

Jennings [2] in a very interesting and valuable

[1] Jacobs, S. M. "Inbreeding in a Stable Simple Mendelian Population with Special Reference to Cousin Marriage. Proc. Roy. Soc., Vol. 84, B, pp. 23–41, 1911.

[2] Jennings, H. S. "Formulæ for the Results of Inbreeding." *Amer. Nat.*, Vol. XLIII, pp. 693–696, 1914.

paper has given a general formula for the determination of the percentage of homozygotes (which he calls the coefficient of homozygosis) after $n$ generations of inbreeding. His results are as follows:

"Let $x$ = the coefficient of homozygosis,
$n$ = the number of inbred generations (the number of times successive brother by sister mating has occurred),
$f_1, f_2, f_3$, etc., = the successive terms of the Fibonacci series (thus $f_1 = 0, f_2 = 1$, etc.).

Then the formula for the coefficient of homozygosis is:

$$x = \frac{2^{n-1} + f_1 \cdot 2^{n-2} + f_2 \cdot 2^{n-3} \cdot \ldots \text{ etc.}}{2^n}$$

(The terms in the numerator are continued until the exponent of 2 becomes 0)."

It is clear that for further analysis of the problem of inbreeding there will need to be much more comprehensive work done upon this second phase, *i.e.* the theoretical Mendelian consequences of the operation of inbreeding. One of the chief results of the present study, in the writer's opinion, is to bring out clearly the importance of a problem which has not yet been considered at all in any of the discussions of the subject which have hitherto appeared. This problem may be put in the following way:

What is the mathematical probability that the

two individuals, $a$ and $b$, which mated together produce the individual $x$, are of the same zygotic constitution in respect of any one or more characters, when they have 1, 2, 3, 4, . . . *m common ancestors* in the $n$th ancestral generation?

This I believe to be the crucial outstanding problem of Mendelian mathematics in relation to inbreeding. Its solution ought to be in principle simple, if somewhat tedious in the carrying out. As has already been pointed out (p. 145, *supra*), it seems likely on *a priori* grounds that this probability will be found always to bear a definite relation to the coefficients of relationship. If this be true, it will be of great help practically in studying inbreeding, since it is always a simple matter to determine coefficients of relationship.

Finally, to summarize briefly this rather extended discussion of the logical aspects of the problem of inbreeding, it may be said that in this paper has been presented, first, a general method of measuring the intensity or degree of the inbreeding practiced in any particular case. The method proposed is shown to be perfectly general. It is based on no assumption whatever as to the nature of the hereditary process. On the contrary, it is founded on the most completely logical and comprehensive definition of the concept of inbreeding that it seems possible to formulate. This is, in simplest form, that the fundamental objective criterion which distinguishes an inbred

individual from one not inbred is that the former has fewer different ancestors than the latter. It is believed that the proposed coefficients of inbreeding may be made extremely useful in studies of the problem of the effect of inbreeding, whether in relation to its purely theoretical aspects, or in the practical fields of stock breeding and eugenics.

In the second place, attention is called to the fact that inbreeding of considerable degree may be brought about in the entire absence of any kinship between the two individuals bred together, and there is described a method of separately measuring what proportion of the observed inbreeding in a particular case is due to kinship of the parents, and what to earlier ancestral reduplication. A proposed coefficient of relationship is described, and its application illustrated by concrete cases.

In a final section it is shown that logically the general problem of inbreeding must be considered as composed of three distinct elements, viz.:

1. Inbreeding itself — a system of mating with definite mathematical characteristics.

2. The genetic consequences of inbreeding.

3. The physiological consequences of inbreeding. A discussion of the second of these elements, the first two sections of the paper having been devoted to a consideration of the first, reviewed the meager work so far done on the theoretical side of the problem. No attempt is made here to discuss in any way the third element of the problem.

# CHAPTER V

## GENETICS AND BREEDING[1]

One of the primary purposes for which the American Breeders' Association was founded was to bring together on a common ground those who were approaching the problem of the improvement of plants and animals by breeding, on the one hand, from the side of practical breeding, and, on the other hand, from the side of the scientific study of heredity. One of these groups stands as the representative of the art or craft of breeding, and the other as the representative of the science of genetics. That each of these two bodies of men has something to learn from the other there is no doubt. Even with the continued and prosperous existence of such an association as this it is certain that actually there is far from being anything like as extensive a mutual interchange of knowledge and opinion between science and practice in breeding as would appear from every point of view to be desirable.

[1] Address of the retiring chairman of the Animal Section of the American Breeders' Association at its Columbia, S.C., meeting in January, 1913. Originally printed, in slightly different form, in *Science*, Vol. XXXVII, pp. 539–546, 1913.

It will have been perceived by all who have followed my remarks so far that they have been commonplace to the point of utter banality. They constitute a thoroughly bromidic introduction to a conventional treatment of that time-stained and battle-scarred old theme of compulsory oratory, the relation of science and practice. Every one can foresee, with a moment's reflection, just what ought to come next, and next, and on to the end. At the outstart should be set forth the great achievements of the science of genetics; then should be sketched the tremendous possibilities thus opened out to the practical breeder, who in the near future will be able to soar from this scientific foundation to realms of wealth and power in the community hitherto possible only to the predatory classes; nevertheless, in a meek and humble spirit of gratitude engendered by the blessings which have been poured at his feet, he in turn contributes to the great cause of science by placing at the disposal of the geneticist the wonderful stores of experience he has accumulated; at the end should come an impassioned plea for "getting together" for the good of agriculture, humanity, and sundry other things, which should, if well done, so titillate the emotions as to send everybody home uplifted, and, in general, determined to lead a better life.

I have sketched this little picture, which, if necessarily impressionistic, is essentially true, only to bring into sharp relief the intellectual

junction point, at which we shall alight and change cars. Just because there has been so much perfervid oratory, loose thinking, and cheap advertising of the achievements of men and institutions based on the ideal or assumed mutual interrelationship of the science of genetics and the breeders' art, it seems worth while to make a careful objective analysis of the actually existing relations between these adjoining fields of human endeavor. Such an analysis will be attempted in what follows. Specifically the question to which attention is invited is: What essential and fundamental contributions has genetics made to the *practice* of the breeders' art? Or, to put the matter in another way, what particular things does the most highly successful practical animal breeder do now which he did not do, or performed differently, before Mendelism was rediscovered or Darwin wrote?

It is generally agreed that during the past fifteen years there has been a great advance in our knowledge of the fundamental laws of heredity. Indeed, it may fairly be said that more has been gained in this regard within this period than in the entire previous history of this field of knowledge. The new method of investigating heredity which was given by Mendel's work has for the first time made a real analysis of genetic phenomena possible. It was a truly imposing array of organisms and characters which Major Hurst was

able to list at the meeting in commemoration of Mendel at Brünn, as comprising those attributes of organisms about the inheritance of which something *definite* is *known*.

There is a very widespread assumption that coincident with this advance in our knowledge of the fundamental laws of inheritance there has been an equal advance in the practical art of breeding. This has perhaps resulted from the somewhat overenthusiastic prophecies of the early Mendelian workers. Many will remember the glittering possibilities set forth to the practical breeders in the early meetings of this association. They were told in effect that at last the key to the genetic riddle had been found; that by the application of these simple Mendelian laws existing races of animals could be brought up to desired ideals with more certainty and dispatch than had hitherto been possible, and that new races could be created which would surpass in usefulness, anything now existing. There was, of course, an element of truth in all this. But it raised unwarranted hopes in the minds of many laymen. The apparent failure of these prophecies to be realized has probably done real harm to the cause of science in the minds of some practical men — representatives of the class to which in last analysis science must look, for its material support — and very generally has led animal breeders to underrate the real value of Mendelian investigations.

It is permissible to think that the fundamental error involved was in the assumption we are all inclined to make that any distinct advance in science necessarily means an equally marked and immediate advance in the practice of the associated art or craft. It is extremely difficult for the man of the laboratory or the study, as he takes a broad view of the history of the industrial arts, and sees that great progress there has rested upon fundamental scientific discoveries, to realize that the art of breeding differs essentially in this respect from the industrial arts. The breeding of animals by man for more or less definite purposes goes back to prehistoric times. Practically as soon as primitive man began the domestication of animals he must perforce have begun, in greater or less degree, to control their breeding. Having started thus early, the craft of breeding had attained a relatively high degree of development centuries before any attempt was made to formulate the scientific principles of genetics. As an example may be mentioned the breeding of horses in England. It is customary to think of "stallion laws," aimed at the improvement of the horses of a state, as very modern and American, and an indication of the influence of the science of breeding on the practical craft. But three hundred and seventy-odd years ago, in the reign of Henry VIII, there was a "bill for the breed of horses," which in preamble stated that:

"Forasmuch as the generation and breed of good strong horses within this realm extendeth not only to a great help and defence of the same, but also is a great commodity and profit to the inhabitants thereof, which is now much decayed and diminished by reason that, in forests, chases, moors and waste grounds within this realm, little stoned horses and nags of small stature and of little value be not only suffered to pasture thereupon, but also to cover mares feeding there, whereof cometh in manner no profit or commodity."

In order to prevent the multiplication of poor specimens section 2 of this law provided that no uncastrated stallion two years or more old which was under 15 "handfulls" high should be allowed to graze on common or waste land in certain counties. Further, it was provided in section 6 that all forests, chases, commons, etc., were to be "driven" at a stated time in the year (just preceding Michaelmas day) and all horses, mares, and colts which were not of good quality, or did not promise to become or to produce serviceable animals, were to be killed.

The fact is that the practice of the art of animal breeding, so far from languishing, for want of instruction from the science of genetics is actually immeasurably in advance of that science. The geneticist who is disposed to think otherwise should visit a great horse, or cattle, or even poultry show, and then permit himself to consider candidly the

question whether with all his science he could himself breed, or tell any one else how to produce, *finer* specimens than he will see there. Yet by hypothesis that is exactly what he ought to be able to do, if genetics is to set up as a teacher and guide to the best practical methods of live-stock breeding.

It is capable of abundant historical proof that many years ago, before the beginning of the world movement towards agricultural education, experimentation and the grounding of a science of agriculture in general, there were in existence individual animals (even flocks and herds), and strains of seeds of farm crops which were probably intrinsically as fine, as productive, and generally as excellent as any that we know to-day. Given as intelligent care and feeding as our prize-winning animals and plants now get, there is every reason to believe that they would have equaled or surpassed our finest specimens of to-day. Some specific examples may be cited. Mr. George A. Scott,[1] of Nashville, Tenn., had in 1863 "a common scrub cow" which produced in one year 1447½ gallons of milk. Taking the weight of one quart of milk at 2.15 lbs. as sufficiently close for practical purposes, this gives a record of 12,448.5 lbs. of milk for the year. This is a respectable figure even for present standards. Going back half a century earlier, we have the record of a

[1] *The Cultivator and Country Gentleman*, Vol. 28, p. 401, 1866.

Sussex cow:[1] "a cow not of either of the highest improved English breeds — long horns or short horns; but of the proper old Sussex breed." The following record is of her production in five successive years beginning in 1805. I have transposed quarts to pounds by the use of the factor given above:

| | Weeks in Milk | Lbs. of Milk | Lbs. of Butter |
|---|---|---|---|
| First year . . . . | 48 | 10,580.2 | 540 |
| Second year . . . . | 45½ | 8,894.6 | 450 |
| Third year . . . . | 51½ | 12,366.8 | 675 |
| Fourth year . . . . | 42½ | 9,070.9 | 466 |
| Fifth year . . . . | 48 | 11,543.4 | 594 |

Facts of the same sort are at hand for crops. Justin Ely, Esq., of West Springfield, Mass., in 1816, raised 50 bushels of wheat to the acre. Colonel James Valentine, of Hopkinton, raised 128 bushels of "Indian corn" to the acre. Payson Williams, Esq., of Fitchburg, raised 614 bushels of potatoes to the acre, and James Whitton, Esq., of Lee, raised 85 bushels of oats to the acre. The average yield of oats to-day is approximately 36 bushels to the acre. The Maine Agricultural Experiment Station, in its tests of the best commercial varieties of oats procurable in this country and Europe,

[1] *Massachusetts Agricultural Repository and Journal*, Vol. IV, No. 4. Cf. also *New England Farmer*, Vol. III, p. 305, 1825.

has, up to the time of writing, never been able to obtain a yield per acre of more than 76 bushels.[1]

I have elsewhere discussed records of egg production in poultry in this connection. From 1836 there is an authenic record of crested Polish fowls producing an average of 175 eggs each per year. This was long before the trap nest had been discovered.

Too much stress, of course, should not be laid on such examples as these. They do not indicate that there has been no advance made by the breeder in the qualities of domesticated animals and plants during the last century. The *average* quality of live stock and of crop plants is constantly improving, not only as a result of breeding but also because of better and more widely disseminated knowledge of how to provide the food and environmental conditions best suited to bring to full expression the potential hereditary capabilities [2] of the individual. I think that such records, however, do fairly indicate that in the practice of the art of breeding there has been no such marked fundamental advance in recent years as there has been in the science of genetics. By empirical methods man has been steadily im-

[1] In the season of 1914 this was exceeded.

[2] Consider in this connection the practices of the real expert in making world's records for milk and butter fat production in the seven- and thirty-day advanced registry tests of the Holstein-Friesian breed.

proving the quality of live stock for centuries past, and long ago a *relatively* high level was reached by the most skillful breeders.

Furthermore, in this same connection, the fact must not be lost sight of, that in the practice of the breeder's art many of the most brilliant successes are purely accidental, in the sense that the superior individual often appears quite without relation to the breeder's conscious or planned efforts. A few examples will suffice to illustrate this point. Mr. Oscar Tretsven of the Montana Agricultural College, has reported recently [1] the particulars in regard to the performance of a *grade* Jersey cow purchased in Minnesota. No particulars of its breeding were given. It was just a "grade," that, like Topsy, "growed." [2] However, her record for a year was 16,286.1 lbs. of milk, 844.8 lbs. of fat (= 1056 lbs. of 80 % butter). In a seven-day test she produced 450.2 lbs. of milk and 21.245 lbs. fat. Her yearly record, at the time it was made, put her fifth in the list of highest producing (world's record) cows of the Jersey

[1] "*Hoard's Dairyman*, Vol. XLIII, p. 695, May 31, 1912.

[2] In a letter of July 14, 1914, Professor R. F. Miller very kindly gives me the following information about this cow: "I may say that we do not know anything about her breeding. She was simply bought in Minnesota from a Polish farmer with a lot of other grade milch cows. The man we bought her of was not making dairying a specialty and had probably not bred his stock very strictly in that direction. We consider her to be a grade Jersey, although she shows some Guernsey blood and a Shorthorn frame."

breed. The only four records then exceeding that of this grade cow were those of the famous Jersey cows, Jacoba Irene, Sophie 19th of Hood Farm, Olga's 4th Pride, and Adelaide of Beechland.

Another similar example is found in a recent report of the performance of a scrub cow in Washington, D.C.[1] Since this report is very brief, it may be quoted in full.

"An old black scrub cow No. 131 in a local government herd was last fresh Oct. 6, 1909, and has since been sterile, due to old age (probably 15 to 20 years old). From this date to Aug. 1, 1913, she has made 33,066 lbs. of milk."

This is a remarkable record, and it is quite clear that neither the science of the geneticist nor the art of the breeder had anything to do with the producing of this old black scrub.

Purely empirical methods are wasteful and slow in operation, but they may attain excellent results. When they are successful it is obviously because at just that point the practice was, by chance, in exact conformity with the underlying principle or law concerned. More generally it may be said that all progressive success of empirical methods depends on a gradual elimination of those operations or practices which do not accord with basic natural laws. In the consideration of the science and practice of breeding this has some-

[1] *Hoard's Dairyman*, Vol. XLVI, p. 175, Sept. 12, 1913.

times been forgotten. It is difficult to remember always that a law of nature may be presumed to have been in operation before its discovery. If Mendel's law represents a real and fundamental law of nature, as certainly appears to be the case in the light of present evidence, it is quite certain that it did not begin operation in A.D. 1900. Whatever of success has been attained during centuries past in the breeding of improved strains of animals and plants must have been attained by methods and practices which were not violently in discord with Mendelian principles. A nomad Arab may never have heard of the principle of segregation, but none the less he had to reckon with the phenomenon in breeding his horses.

Looking at the matter in this way, the reason is clear why the rediscovery of Mendel's work and the brilliant genetic researches which have followed did not and could not have had any profound revolutionary effect on the *practice* of the animal breeders' art. By years — even centuries — of "trial and error" methods, breeding practice has been brought into rather close conformity with the basic laws of heredity. The discovery of some of these laws by the geneticist could not radically change the breeder's way of attaining results.

What then has the rapidly developing science of genetics done for the breeder and what can it do? Still looking at the matter from the standpoint of

the practical animal breeder, it must be agreed, I think, that the chief contribution of recent discoveries in the field of inheritance is that they have brought to light and fairly established certain general principles which enable him in greatly increased measure to understand and interpret his methods and his results.[1] This may seem too mild a statement of the practical value of genetic science to the animal breeder. It undeniably does lack the grandeur of the vision sometimes opened out by the extension lecturer in his zeal to inspire the farmers to better things, and at the same time pave the way to increased appropriations for his institution. But to help one to understand and to interpret is, after all, no mean achievement. It signifies that, with much economy of effort, the successful breeder may dispense with the merely trivial and unessential in his empirical methods, and more directly and uniformly attain the same or a greater measure of success than before. To his less successful brother and the beginner, it means a surer and more rapid guide than the old tradition based on empiricism. It is certain that

[1] This is of course to be understood as a general statement. There are now a few specific instances, and in time there will be more, where the geneticist has been able to show the breeder precisely how to attain a particular result in breeding commercially for a particular quality, which result he had only hitherto been able to obtain by chance. In no such case, however, so far as I am aware, has the new method been so essentially different from former practice as to be fairly regarded as "revolutionary."

the young man starting out to-day to be a breeder of fine cattle, of fine horses, of fine chickens, is likely to attain his goal somewhat sooner if he thoroughly understands the meaning of those laws of inheritance associated with the name of Mendel.

The most important general principles which the scientific study of genetics has firmly grounded are, it seems to me, these:

(*a*) That the fundamental basis of all inheritance is to be found in the germinal constitution of the individual rather than in the body or soma. Those qualities alone are inherited, which are innate in the germ cells, the ova and the spermatozoa. Here only can the breeder find the means with which to accomplish his ends. However interesting theoretically may be those rare and still doubtful cases in which extraordinary influences acting upon the body under the controlled and special conditions of the laboratory may perhaps influence the germ cells through the soma, they have no bearing on the practical conduct of the breeders' craft. Genetics has demonstrated that he may cast aside, for once and all, that mass of tradition and superstition which assumed that influences specifically affecting the body will specifically modify subsequent generations. Has not genetics done breeding a service of great value in freeing it of the sinister influence of "telegony," "saturation," "maternal impressions" and similar sorts of nonsense?

(*b*) That specific characters or groups of characters, in the great majority of cases and perhaps all, are inherited as discrete and definite units. If one mates a pea-combed fowl with a single-combed fowl, all the offspring will have pea-combs. This result occurs whether the pea-combed parent is a Game or a Brahma; whether it is a male or a female; whether it is a strong, vigorous individual, or the sickliest, weakest scrub in the flock. In other words, the *kind of a bird* it is whose germ cells carry the potentiality to make pea-combs develop in the offspring, so far as we now know, has nothing to do with the *specific* result (*i.e.*, the production of a *pea*-comb, rather than a single, a rose, or any other kind). Comb form is inherited as a discrete unit largely, if not completely, uninfluenced by the individual's other attributes. This discovery that many characters are inherited as separate units — and no principle of genetics is more firmly grounded than this — gives the breeder a totally new concept of the meaning of "purity" of blood in breeding. We see now that properly (*i.e.*, biologically) one can only speak of an animal as being "pure-bred" when he specifies the particular *character* to which he refers. A chick may be the veriest mongrel in all other respects and yet carry in the germ cells only that potentiality in respect of comb form which leads to the development of a pea-comb. Then however much of a mongrel it may be in

respect of all other characters, it is "pure" and "pure-bred" so far as concerns comb. Is it not a contribution of moment to the breeder to have demonstrated that in his breeding operations he may safely and surely deal with individual characters, and groups of correlated characters as units?

(*c*) That in a very great range of cases, perhaps in all — the number of known cases daily grows larger — the Mendelian law of segregation and recombination of characters operates. In the formation of the germ cells of an individual there is a sorting out or segregation of the hereditary characteristics contributed by the father and the mother and a readjustment of these into all of the combinations, both old and new, which are mathematically possible. What may be the precise cellular mechanism or basis of this wonderful process is not altogether certain, but the phenomenon itself is as certain as the phenomenon of gravitation. It operates as well in regard to the minutest heritable differences in the pedigreed specimens of the same sub-breed as in the wide differences of true hybridization. Properly understood, it enables the breeder to interpret and weigh the results of his breeding operations, and so intelligently to plan the next steps with a certainty and precision hitherto unattainable. Is not this a real contribution of science to practice?

(*d*) That the germinal bases of heritable unit characters can be changed or altered in any respect,

only with the greatest difficulty, if at all. It is, I believe, fair to say that there is at present no critical, unchallenged evidence that any alteration can be produced. This matter has recently been discussed in a most able manner by East.[1] The weight of evidence at present indicates that selection does not act in the manner it was long supposed to, in accordance with Darwin's interpretation. It appears that selection, however stringent or long continued, is powerless to alter in any way the original potentialities of the germinal basis of a unit character. Selection appears to be essentially a process of sorting out from a mixture of heritable variations what is already there, and not a germinally creative or germinally additive process.

So far this discussion has been approached from the standpoint solely of animal breeding. It is perhaps allowable, even before this animal section, to digress for a little and discuss plant breeding. The ultimate objective point of the animal breeder is the same as that of the plant breeder; namely, the greatest possible improvement of animals and plants and their adaptation to the needs of man. The practical method of working towards this goal is, however, somewhat different in the two fields. The animal breeder almost exclusively works towards the amelioration of existing fixed and "pure" breeds. Especially among the larger

[1] *American Naturalist*, 1912.

domestic animals such a thing as a *new* breed is brought forward by the breeder only on very rare occasions. Almost all of our existing breeds of horses, cattle, sheep, and swine have long histories as "pure breeds," and no new ones are being added now. With smaller animals such as poultry the case is of course somewhat different. There we have no registered pedigrees and, with some difficulty, new breeds may be launched.

The plant breeder, on the other hand, makes nearly all of his improvements by the production of new varieties. This he does either by hybridization, actually building up a new type, or by isolation of superior pure-breeding forms from already existing mixtures. He is not hampered by a body of tradition that only the "pure-bred" is of any particular value. Almost if not quite every one of the most valuable strains of agricultural plants to-day carries the "bar sinister." To the animal breeder they would be "grades" or "crosses" however gametically pure, and only with the greatest difficulty would ever have gained a chance to show their worth.

No one would deny that the systems of registry for live stock and the exploitation of the "pure-bred" have been of great value in the development of the animal industry of the world. They certainly have; and every day the economic importance of the system becomes greater, for obvious reasons. All systems of pedigree registra-

tion operate economically precisely like a monopoly. As such a plan of developing the live-stock industry of a country grows, the more difficult does it become for a *new* creation of the breeder to get a foothold. If it is new, it is by definition not "pure-bred," because if it were "pure-bred," it must belong to one or another of the established breeds. But anything not "pure-bred" has no recognized standing, or market value. Without regard to the merits of the individual the mere fact of pedigree registration adds a definite and not inconsiderable amount to the monetary value of an animal. In last analysis this fact is to-day one of the strongest arguments which can be made to the farmer in favor of keeping "pure-bred" animals.

What has just been said is not intended in any way to criticize, or belittle the importance and value of the "pure-bred" registry system of developing the live-stock industry of the world. I merely wish to point out that when he adopted the system, the animal breeder took upon himself along with the advantages certain very real restrictions to the freedom of his breeding operations, which the plant breeder has escaped. The animal-breeding industry of the world has developed as a system of pedigreed aristocracy. The plant-breeding industry is developing as a democracy. The "social position" of a horse or a cow is primarily determined on the basis of whether it had a grand-

father or not. A variety of oats takes its place in the world by virtue of its own inherent qualities, with no questions asked about forebears or the orthodoxy of their marital relations. Both aristocracies and democracies have their advantages and their disadvantages as social systems. These merits and defects are just as real and effective in their operation whether the ultimate vital unit of the system be a man, a cow, or an oat plant.

Owing to the essentially different conditions and methods of work which obtain in plant breeding, this field is able to reap more direct benefits of a practical character from the advances which have been made in the science of genetics, than in animal breeding. In the creation of new races by hybridization the plant breeder can and does take Mendelian principles as a direct and immediate guide. He has made Mendelism a working tool of his craft.

To conclude: What I have tried to do in this paper is to discuss the relation between the science of genetics and the practical art of breeding as they actually have developed and now exist. Attention has been directed to the obvious fact that animal breeding has, without the aid of genetic science, attained an extremely high level of achievement. Empirical methods can only have been successful when they were fundamentally in accord with natural laws, and it is therefore not to be considered surprising that the recent discoveries of

world-old genetic laws have not radically modified the successful animal breeders' methods. In pointing out that a scientifically trained geneticist is not as yet an absolutely indispensable necessity on a successful animal-breeding farm I have no thought or desire to belittle the importance of the science of genetics. My zeal and enthusiasm for the advance of knowledge in this field know no bounds. This attitude, however, furnishes no reason that the geneticist should delude himself, or by rash statements hold out false hopes to the breeder, as to the immediate practical importance of some of the recent developments in the science of genetics. All knowledge is potentially useful, but the fundamental reason for undertaking and encouraging research in genetics, or anything else, is not because what one gets may be immediately useful, but because it is *knowledge*.

# INDEX


Abstract qualities of groups, 80.
Accuracy of statistical constants, 53.
Adami, 41.
American Breeders' Association, 159.
Ancestral elimination, 114.
Ancestral generations, designation of, 141.
Ancestral inheritance, law of, 11, 13, 49, 66, 68.
Anthropology, 42.
Association, degrees of, 92.
Assortative mating, 10.
Avuncular matings, 126 ff.

Bailey, 27.
Bateson, 21, 27, 32.
Beard illustration, 13.
Biochemical method in inheritance, 40.
Biological implications of biometry, 63.
Biometric constants, 57, 80.
Biometric method in genetics, 9 ff; compared with Mendelian, 19.
Biometry, 9, 62 ff.
Biometry a descriptive method, 12, 54, 62, 91.
Births of children, illustration, 92.
Blood transfusion, 9.
Boring, 38.
Boveri, 25.
Brooks, 36.
Brother × sister mating, 111 ff.
Bunsow, 102.
*Bursa bursa-pastoris*, 17.

Card drawing illustration, 36.
Carrel, 52.
Castle, 17.
Causation, 87.
Center of group, 80.
*Ceratophyllum*, 58.
Chapeaurouge, 102.
Chromosomes in inheritance, 25, 26, 31, 33–35.
Clerk-Maxwell, 85.
Coefficient of inbreeding, 110, 113, 115, 119, 120, 125, 126, 130, 132, 140, 145.
Coefficient of relationship, 133 ff., 145; defined, 143.
Comb form, inheritance of, 173.
Composition of groups, 79.
Conklin, 7, 36.
Contingency, degree of, 82.
Cook, 105.
Corn, selection experiments with, 16.
Corn, yield of, 166.
Correlation coefficients, 15, 61, 69.
Counting, 79, 130.
Cousin mating, 121 ff.
Culture *in vitro*, 5.
Curve fitting, logic of, 21, 28.
Curves of inbreeding, 110, 116, 139.
Cytological method in genetics, 24 ff.
Cytoplasm in inheritance, 34.

Darwin, 9, 161, 175.
Description as philosophical category, 13.
Description embryology, 37.
Development, 5.
Differentiation, 5.
Diversity, individual, 81.
Doncaster, 34.
Double cousins, 121, 123, 125.
*Drosophila*, 31.

East, 16, 145, 149, 175.
Eggs, yield of, 167.
Ely, 166.
Embryological method in genetics, 35 ff.
Enriques, 14.
*Entwicklungsmechanik*, 35.
Expectation, 30.
Experimental method, 13, 18, 20, 62.

Factors, multiple, 21, 27.
Fallacy of qualitative inference, 28.
Feathers, 8, 32.
Fecundity, 17, 67.
Fiddler crab illustration, 96.
First-cousin × first-cousin mating, 121 ff.
Fish, 150.
Fluctuating variation, 48.
Free generations, 108.

Galton, 9, 10, 11, 17, 42, 43, 49.
Gametogenesis, 4, 12, 24.
Generation interval in cattle, 128.
Genetics and breeding, 159 ff.
Genetic bearing of inbreeding results, 147 ff.
Genetics, practical results of, 171 ff.
Genotype, 65.
Germ cell, 4, 27.
Group description, 56.
Gurwitsch, 36.
Guyer, 41.

Hagedoorn, A. C., 17.
Hagedoorn, A. L., 17.
Hatai, 22.
Hatschek, 41.
Hayes, 16, 145, 149.
Henry VIII, 163.
Heredity, problem of, 2 ff.; categories of, 3.
Heterogeneity measured by correlation coefficient, 69.
Hoesch, 102.
Holstein cattle, 144, 167.
Homozygosis, coefficient of, 155.
Homozygotes, proportion of, in inbreeding, 149 ff.
Horse, 107, 163.
Hurst, 161.

Inbreeding, 101 ff.
Inbreeding, calculation of coefficients of, 111 ff; curve of, 110, 116.
Inbreeding, definition of, 106.
Individual, importance of in biological analysis, 48.
Intensity of inheritance, 65.

Jacobs, 155.
Jennings, 6, 7, 13, 58, 67, 69, 89, 149, 155.
Jersey cattle, 128, 129, 136 ff., 144, 169.
Johannsen, 65, 67, 69.

Kelvin, 50.
Kinds of knowledge given by statistics, 79.

King Melia Rioter 14th, 135 ff.
Kossel, 41.

Law of ancestral inheritance, 11, 13, 49, 66, 68; biological error in, 68.
Lehndorff, 102, 107, 108.
Like parts, 58.
Line breeding, 105.
Line of descent in calculating inbreeding, 130.
Linkage phenomena in inheritance, 31.
Location of hereditary determiners, 32.
Low, 102, 108.

Maternal impressions, 172.
Mendel, 161.
Mendelian mathematics, 21.
Mendelian method, 19 ff.
Mendelism, limitations of, 23.
Mendelism, practical bearings of, 162.
Merz, 9, 76.
Milk production, 165, 166, 168, 169.
Miller, 168.
Minot, 34.
Morgan, 31, 32.
Morphogenesis, 59.
Mule illustration, 90.
Multiple factor hypothesis, 21.

Nerve-muscle physiology, 43.
Nilsson-Ehle, 21.

Oats, yield of, 166.
von Oettingen, 102, 107.

Pangenesis, 9.
*Paramecium*, 6, 58, 67.
Parent × offspring mating, 117 ff.
Pawlow, 52.
Pearl, 17, 29, 38, 53, 58, 66, 67, 150.
Pearson, 9, 10, 11, 13, 21, 22, 42, 43, 49, 64.
Pedigree as unit of analysis, 15.
Pedigree table, brother × sister (hypothetical), 112; parent × offspring (hypothetical), 118; single cousins (hypothetical), 122; double cousins (hypothetical), 123; uncle × niece (hypothetical), 127; Bess Weaver, 129; King Melia Rioter 14th, 142.
Pflüger, 41.
Phænotype, 65.
Phenomenal indeterminism, 86.
Phillips, 17.
Plant and animal breeding contrasted, 175.
Postembryonic development, 37.
Potatoes, yield of, 166.
Poultry, 8, 17, 38, 66, 67, 167, 173.
Practice of breeding, 161.
Prediction of future events, 94.
Probable error concept, 50, 59, 96.
Probability, science of, 77.
Punnett, 27, 32.
"Purity of blood," 173.

Qualitative and quantitative aspects of characters, 67.
Quantification of biology, 50.
Quetelet, 42.

Rats, selection experiments with, 17.
Redfield, 128.
Reduplication hypothesis, 27, 30.
Regeneration, 5.
Registration of live stock, 176.

Relationship coefficients, 133 ff.; defined, 143, 145.
Resemblance between individuals, 3, 10, 15, 64.
Ritter, 55.
Royce, 14, 74, 85, 88.

Saturation, 172.
Scott, 165.
Segregation, 25, 45, 174.
Selection, 15, 48, 175.
Sex chromosomes, 26.
Shull, 17.
Single cousins, 121, 122, 125.
Skewness, 81.
Somatogenesis, 5, 12, 35.
Specificity, germinal, 8, 12.
Specificity, somatogenic, 5, 8.
Spillman, 24.
Statistical knowledge, 73 ff.; analysis of, 79; exactness of, 81.
Statistical method, possibilities of, 83.
Statistical probabilities, 92.
Strang, 102.
Surface, 16, 58, 66.
Symmetry of distributions, 81.
Systematic zoölogy, 56.

Telegony, 172.
Tobacco, selection experiments with, 16.
Tretsven, 168.
Trow, 27.

Uncle × niece mating, 126 ff.
Unit characters, 173.

Valentine, 166.
Variability, 8.
Verworn, 41.

Weldon, 42.
Wheat, yield of, 166.
Whitten, 166.
Williams, 166.
Wilson, 33.

Yule, 22, 71.

Printed in the United States of America.

The following pages contain advertisements of a few of the Macmillan books on kindred subjects

By S. HERBERT, M.D. (Vienna), M.R.C.S. (Eng.), L.R.C.P. (Lond.)

# The First Principles of Heredity

*Cloth, 199 pages, Ill., 8vo, $2.00*

The purpose of this book is to supply in a simple and yet scientific manner all that may be desirable for the average student to know about Heredity and related questions, without at the same time assuming any previous knowledge of the subject on the reader's part.

# The First Principles of Evolution

By S. HERBERT

*Cloth, 8vo, 346 pages, containing 90 illustrations and tables, $1.60*

Though there are hosts of books dealing with Evolution, they are either too compendious and specialized, or, if intended for the average reader, too limited in their treatment of the subject. In a simple, yet scientific, manner, the author here presents the problem of Evolution comprehensively in all its aspects.

CONTENTS


Introduction — Evolution in General.
Section I — Inorganic Evolution.
The Evolution of Matter.
Section II — Organic Evolution.
Part I — The Facts of Evolution.
Morphology.
Embryology.
Classification.
Palæontology.
Geographical Distribution
Part II — Theories of Evolution
Section III — Superorganic Evolution.
Social Evolution.
Conclusion — The Formula of Evolution.
The Philosophy of Change.


---


THE MACMILLAN COMPANY

**Publishers** **64–66 Fifth Avenue** **New York**

# Societal Evolution: A Study of the Evolutionary Basis of the Science of Society

By ALBERT GALLOWAY KELLER

Professor of the Science of Society in Yale University. New York, 1915.

*338 pages, 12mo, $1.50*

The author shows that the evolutionary formula of Darwin, the terms variation, selection, etc., can be carried over to the social field without resting any weight on analogy; and thus there can be given to these terms, which are now used indiscriminately, in social writings of all kinds, something rather more sharp in the way of connotation. He shows the nature of variation when it is in the social field; how social selection is related to, and different from, natural selection; how social transmission (tradition) is performed, and what it does; and how social institutions exhibit adaptation to environment, natural and artificialized. The main thought of the book lies in the discussion of rational selection, how and how far possible. Adaptation shows that any institution — settled institution — is justifiable in the setting of its time, as an adaptation.

As one of the foremost sociologists of the day has said, it is "a serious and important work. Professor Keller carries on the interpretation of society begun by the late Professor William G. Sumner, and greatly adds to the value of Sumner's exposition by this rounding out. As scientific work, it is thoroughly good throughout, sober, well-buttressed, and keenly intelligent at every point. Every student of sociology will welcome it, and it is sure to hold an important place for a good while to come. Keller writes a straightforward, strong, and clear style. All in all, the work stands out as greatly superior to the run of sociological writings for four or five years past."

---

THE MACMILLAN COMPANY

**Publishers** **64-66 Fifth Avenue** **New York**

# Diseases of Poultry: Etiology, Diagnosis, Treatment, and Prevention

By RAYMOND PEARL, FRANK M. SURFACE, and MAYNIE R. CURTIS

*Illustrated, cloth, 12mo*

This book gives a comprehensive survey of the present state of knowledge regarding the diseases of poultry. The point of view of the practical poultry raiser is emphasized throughout. To the average man the chief points of interest about disease are three: — its recognition, its cure, if such be possible, and its prevention in the future. These are the matters on which stress is laid in this volume. Entirely original and outstanding features of the treatise are: (*a*) the discussion of hygiene and sanitation, (*b*) the tables for aiding in the diagnosis of disease both from external symptoms and post-mortem findings, (*c*) the thorough analysis of the literature of the subject, (*d*) the wide first-hand experience of the authors with the subject. Nearly every disease mentioned has been observed by the authors directly in their work at the Maine Agricultural Experiment Station.

# The Germ-Cell Cycle in Animals

By ROBERT W. HEGNER

Assistant Professor of Zoology in the University of Michigan, author of "An Introduction to Zoology," and "College Zoology." New York, 1914.

*346 pages, ill., 12mo, $1.75*

An excellent survey of a subject which is extremely important to all biologists and a book available for use as a supplementary text in courses on Cellular Biology, Evolution, Heredity, and Genetics. The term "Germ-Cell Cycle" is meant to include all those phenomena concerned with the origin and history of the germ cells from one generation to the next generation. Contrary to the usual custom, the period of the germ-cell cycle which is emphasized in this book is not the maturation of the germ cells, but the segregation of the germ cells in the developing egg and the visible substances (keimbahn-determinants) concerned in this process. While the author's treatment is technical, he has presented the data available in such a clear way as to make them intelligible to those who have not been able to follow in detail the progress of cytology during the past few years.

---

THE MACMILLAN COMPANY

**Publishers** **64–66 Fifth Avenue** **New York**

# Genetics. An Introduction to the Study of Heredity

By HERBERT EUGENE WALTER

Associate Professor of Biology, Brown University

*Cloth, 12mo, $1.50*

In his "Genetics" Professor Walter summarizes the more recent phases of the study of heredity and gives to the non-technical readers a clear introduction to questions that are at present agitating the biological world.

Professor Walter's conception of sexual reproduction is that it is a device for doubling the possible variations in the offspring, by the mingling of two strains of germ plasm. The weight of probability, he concludes, is decidedly against the time-honored belief in the inheritance of acquired characters. Professor Walter also predicts that the key to this whole problem will be furnished by the chemist, and that the final analysis of the matter of the "heritage carriers" will be seen to be chemical rather than morphological in nature. In the practical application of this theory to human conservation or eugenics, it would follow that the only control that a man has over the inheritance of his children is in selecting his wife. Professor Walter holds, if only modifications of the germ plasm can count in inheritance, and if these modifications come wholly from the combination of two germ plasms, then the only method of hereditary influence is in this selection.

"I find that it is a very useful study for an introduction to the subject. Professor Walter has certainly made one of the clearest statements of the matters involved that I have seen, and has made a book which students will find very useful because he keeps everything in such entirely simple and clear outlines, and at the same time he has brought the book up to date."—PROFESSOR LOOMIS of Amherst College.

"I am much pleased with it and congratulate you upon securing so excellent a treatment. It is one of the most readable scientific books I have, and goes unerringly to the fundamentals of our most recent advances in the experimental study of heredity as well as those of the older studies."—PROFESSOR GEORGE H. SHULL, Cold Spring Harbor, Long Island, N. Y.

"There was a decided need for just such a work. The book strikes me as most excellently done."—PROFESSOR H. S. JENNINGS, Johns Hopkins University.

---

THE MACMILLAN COMPANY

**Publishers** **64-66 Fifth Avenue** **New York**